

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
25 January 2007 (25.01.2007)

PCT

(10) International Publication Number  
**WO 2007/011722 A2**

(51) International Patent Classification: **Not classified**

(21) International Application Number:  
PCT/US2006/027354

(22) International Filing Date: 14 July 2006 (14.07.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/699,735 15 July 2005 (15.07.2005) US

(71) Applicant (for all designated States except US): **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02139 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **LIU, David, R.** [US/US]; 1 Fox Run Lane, Lexington, MA 02420 (US). **KANAN, Matthew, William** [US/US]; 44 Langdon Street, Cambridge, MA 02138 (US). **ROZENMAN, Mary, M.** [US/US]; 19A Forest Street, Apt. #24, Cambridge, MA 02140 (US).

(74) Agent: **BAKER, C., Hunter**; Choate, Hall & Stewart LLP, Two International Place, Boston, MA 02110 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

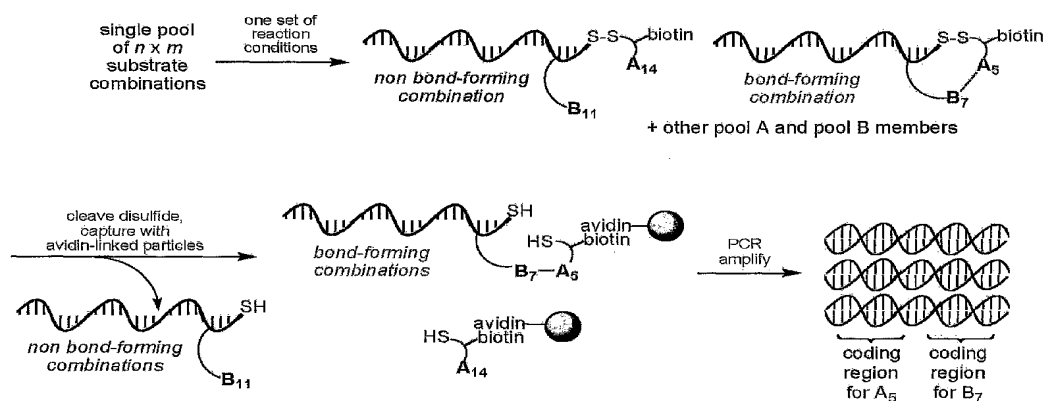
**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: REACTION DISCOVERY SYSTEM

## Selection for Bond-Formation from a Single-Pool Reaction Discovery System



(57) **Abstract:** A novel reaction discovery system that does not depend on DNA duplex formation is provided. The advantages of this system include exploring reactions conditions not possible where DNA hybridization is required. For example, the inventive reaction discovery system allows for reaction conditions using organic solvents, higher temperatures, and water-insoluble reagents, catalysts, and ligands. The invention also provides single-stranded oligonucleotide templates with substrate pairs covalently attached and methods of screening for reaction conditions that result in a direct covalent bond between the substrates. Kits are also provided for practicing this novel reaction discovery system.

## REACTION DISCOVERY SYSTEM

### Related Applications

[0001] The present application claims priority under 35 U.S.C. § 119(e) to U.S. provisional patent application, USSN 60/699,735, filed July 15, 2005, incorporated herein by reference.

### Government Support

[0002] The work described herein was supported, in part, by a grant from the Office of Naval Research (N00014-03-1-0749) and the National Institutes of Health (GM065865). The United States government may have certain rights in the invention.

### Background of the Invention

[0003] Traditional approaches to reaction discovery typically focus on one particular chemical transformation. Predicted precursors for a target structure are chosen as substrates, and then particular reaction conditions are evaluated either manually or in a high-throughput format (Stambuli *et al.* Recent advances in the discovery of organometallic catalysts using high-throughput screening assays. *Curr. Opin. Chem. Biol.* **7**, 420-426 (2003); Reetz, Combinatorial and evolution-based methods in the creation of enantioselective catalysts. *Angew. Chem. Int. Ed.* **40**, 284-310 (2001); Stambuli *et al.* Screening of homogeneous catalysts by fluorescence resonance energy transfer. Identification of catalysts for room-temperature Heck reactions. *J. Am. Chem. Soc.* **123**, 2677-8 (2001); Taylor *et al.* Thermographic selection of effective catalysts from an encoded polymer-bound library. *Science* **280**, 267-70 (1998); Lober *et al.* Palladium-catalyzed hydroamination of 1,3-dienes: a colorimetric assay and enantioselective additions. *J. Am. Chem. Soc.* **123**, 4366-7 (2001); Evans *et al.* Proton-activated fluorescence as a tool for simultaneous screening of combinatorial chemical reactions. *Curr. Opin. Chem. Biol.* **6**, 333-338 (2002); each of which is incorporated herein by reference) for their ability to produce the desired target product. Although this approach is very useful in addressing specific chemical problems, it does not lend itself to the discovery of entirely new chemical reactions. In fact, its focused nature may leave many areas of chemical

reactivity unexplored.

[0004] Recent developments in DNA-templated synthesis suggest that DNA annealing can organize many substrates in a single solution into DNA sequence-programmed pairs. DNA-templated synthesis and *in vitro* selection may, therefore, be used to evaluate many combinations of substrates and conditions for bond-forming reactions (Calderone *et al.* Directing otherwise incompatible reactions in a single solution by using DNA-templated organic synthesis. *Angew. Chem. Int. Ed.* **41**, 4104-8 (2002); Gartner *et al.* The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules. *J. Am. Chem. Soc.* **123**, 6961-3 (2001); Gartner *et al.* Expanding the reaction scope of DNA-templated synthesis. *Angew. Chem. Int. Ed.* **41**, 1796-1800 (2002); Rosenbaum *et al.* Efficient and Sequence-Specific DNA-Templated Polymerization of Peptide Nucleic Acid Aldehydes. *J. Am. Chem. Soc.* **125**, 13924-5 (2003); each of which is incorporated herein by reference). See also published U.S. patent application 2004/018042, published September 16, 2004, which is incorporated herein by reference. Watson-Crick base pairing controls the effective molarities of substrates tethered to DNA strands. Selection for bond formation, amplification by PCR, and DNA array analysis then reveals bond-forming substrate combinations and conditions. The versatility and efficiency of DNA-templated synthesis enables the discovery of reactions between substrates typically thought to be unreactive.

[0005] DNA-templated synthesis has now been used to discover new chemical reactions that are potentially broadly useful in the synthesis of chemical compounds such as pharmaceutical agents, new materials, polymers, catalysts, *etc.* In particular, a DNA-templated reaction discovery system has been used to discover a novel palladium-catalyzed carbon-carbon bond forming reaction. See U.S. Patent Application U.S.S.N. 11/205,493, filed August 17, 2005; Kanan *et al.* "Reaction Discovery Enabled by DNA-Templated Synthesis and *In Vitro* Selection" *Nature* **431**, 545-549 (2004); each of which is incorporated herein by reference. However, the need for DNA hybridization in the reaction discovery system limits the reaction conditions that can be explored. Duplex formation typically requires an aqueous solution with a relatively high salt concentration. Although water has been used extensively as a solvent for organic reactions (Li & Chan, *Organic Reactions in*

*Aqueous Media* John Wiley & Sons, Inc., 1997; incorporated herein by reference), many ligands, catalysts, and reagents are insoluble in water. To access more traditional organic and organometallic chemistry reaction conditions in a selection-based approach to reaction discovery, alternative systems for organizing pairs of substrates in a single solution need to be explored.

### Summary of the Invention

[0006] Any reaction discovery system capable of simultaneously evaluating in a single solution many combinations of substrates for their ability to form new bonds and covalent structures should optimally address the following criteria: (1) the system should organize complex substrate mixtures into discrete pairs that can react (or not react) without affecting the reactivity of the other substrate pairs; (2) the system should include a general method for separating reactive substrate pairs from unreactive pairs; and (3) the reactive substrate pairs should be easily identifiable. Although DNA-templated reaction discovery satisfies each of these criteria, it is limited to exploring reaction conditions that facilitate DNA duplex formation (*e.g.*, aqueous environments, lower temperatures, high salt concentrations).

[0007] The present invention stems from the recognition that many ligands, catalysts, and other reagents used by organic chemists are not soluble in the aqueous, high salt media needed for DNA duplex formation. Therefore, the instant reaction discovery system provides an alternative and improvement to previous reaction discovery systems based on DNA hybridization. The present system does not require duplex formation; instead, the potential substrate pairs are organized on a single nucleic acid (*e.g.*, DNA) strand (*i.e.*, the template). Since the inventive system does not require duplex formation, the reaction condition being explored may also include higher temperatures than possible with the earlier reaction discovery system.

[0008] To eliminate the need for duplex formation, a single pool of nucleic acid templates, each of which is linked to a unique pair of substrates is used in the inventive system. The two substrates are attached to the nucleic acid template molecule in such a way that they can react under suitable conditions. In certain embodiments, one substrate is attached to the 5'-end of the template via a cleavable linker. The other substrate is then attached proximal to the 5' end (*e.g.*, at an internal

modified base). The template molecule includes sequences encoding the identity of each of the substrates attached thereto. Thus, as shown in *Figure 1*, each unique pair of substrates to be tested is linked to a unique template that encodes the identity of the two substrates. Substrates attached in this manner to single nucleic acid template molecule have the same opportunity to react with each other under suitable reaction conditions, similar to substrates attached to two different DNA strands in a DNA duplex.

[0009] Although the invention reaction discovery system does not require nucleic acid hybridization, the solubility of the nucleic acid template molecule is addressed. In certain embodiments, the reaction discovery method is performed at low concentrations (*e.g.*, 0.1-0.0001  $\mu\text{M}$ ) to facilitate the solubility of the template molecules. In other embodiments, an organic solvent-water mixture is used as the solvent. In many cases, ligands, catalysts, and other reagents are insoluble in 100% aqueous solutions but are readily soluble in organic solvent-water mixtures (*e.g.*, mixed aqueous-organic systems that include miscible organic solvents such as acetonitrile, DMF, DMSO, methanol, or dioxane). Furthermore, in certain embodiments, solubility of the template molecule may be enhanced by the use of ammonium salts such as tetraalkylammonium salts.

[0010] In one aspect, pools of template molecules with pairs of substrates attached are dissolved in an organic solvent or an organic solvent-water mixture. *See, e.g., Figure 2*. The solution is then subjected to a particular reaction condition (*e.g.*, temperature, catalyst, reagent, *etc.*). The pool of template molecules is then exposed to conditions that cleave one of the substrates from the template molecule (*e.g.*, a reducing agent to cleave a disulfide bond). If the substrate did not form a direct covalent bond with the other substrate attached to the template molecule, it will be completely free of the template molecule. If the substrate did form a covalent bond under the reaction conditions the pool was subjected to, it will remain attached to the template molecule through the other substrate. In certain embodiments, the cleavable substrate also is attached to biotin; therefore, substrate combinations that have reacted to form a covalent bond remain covalently linked to biotin and can be isolated using a streptavidin resin or streptavidin-linked magnetic particles. The template molecule with substrate combinations that did not form a covalent bond are not linked to biotin

after the cleavage reaction and are therefore washed away. The sequences of the isolated substrate combinations that reacted to form a covalent bond are then optionally amplified (*e.g.*, using PCR) and analyzed to identify the identities of the reacting substrates. Traditional DNA sequence may be used to analyze the results, or DNA microarray technology may be used (see, *e.g.*, *Figures 15-22*). As will be appreciated by one of skill in this art, multiple selections and analyses may be performed in parallel. Therefore, many different reaction conditions may be screened at once. DNA microarray technologies are particularly useful in analyzing matrices of substrates. The present invention also includes any novel chemical reactions that are discovered using the inventive method and system of discovering new chemical reactions.

**[0011]** In another aspect, the invention provides a system for preparing the template molecules with the substrate combinations attached. The inventive reaction discovery system utilizes nucleic acid template molecules attached to two substrates. A modular approach has been developed for preparing a pool of substrate combinations attached to a nucleic acid template molecule. The system typically involves attaching single substrates to an oligonucleotide and then using enzymatic steps to assemble the full-length template as shown in *Figures 4* and *14*. For example, in the first step, primer extension from an overhang adds a sequence encoding one substrate to the 3' end of a sequence encoding another substrate that is already attached to the oligonucleotide at an internal modified base. The resulting template molecule contains one primer binding site and coding regions for both substrates. This template molecule is attached to one substrate of the pair. In the second step, ligation appends an oligonucleotide with a substrate attached to its 5' end via a cleavable biotinylated linker to the 5' end of the template from the first step. The template molecules are prepared individually or are prepared in parallel (*e.g.*, 5-100 or more at a time). In certain embodiments, the resulting full-length template molecule contains two primer binding sites, two coding regions, and the substrate pair to be tested in various reaction conditions. In this way, one substrate is attached at the 5' end, and the other substrate is attached at an internal modified base (*e.g.*, dT). In another embodiment, the system involves labeling an oligonucleotide with two substrates. This addition of substrates to an oligonucleotide are performed using

techniques for modifying oligonucleotides known in the art. The template molecules with the substrates attached and any intermediates thereto are also considered to be within the scope of the invention. In certain embodiments, the inventive template molecule is DNA-based and includes two substrates covalently attached thereto and sequences identifying the attached substrates. The template molecule optionally also includes primer sequences for PCR amplification and/or sequencing.

**[0012]** The present invention also provides kits for practicing the inventive reaction discovery technology. These kits may include possible substrates, template molecules, DNA molecules, nucleotides, nucleotide analogs, primers, buffers, enzymes (*e.g.*, polymerases, ligases, *etc.*), catalysts, reagents, ligands, organic solvents, microarrays, reagents for polyacrylamide gel electrophoresis, columns, resins, affinity reagents, or other materials that would be useful in practicing the present invention. Preferably, the materials are conveniently packaged with instructions for use. The kits may provide enough materials for any number of rounds of screening for new reactions. In certain embodiments, the kit is designed to allow the user to substitute in his or her own substrates, catalysts, solvent systems, or other reagents. In certain particular embodiments, the kit is designed to allow the user to test his or her own reactions conditions.

### Definitions

**[0013]** Definitions of specific functional groups and chemical terms are described in more detail below. For purposes of this invention, the chemical elements are identified in accordance with the Periodic Table of the Elements, CAS version, Handbook of Chemistry and Physics, 75<sup>th</sup> Ed., inside cover, and specific functional groups are generally defined as described therein. Additionally, general principles of organic chemistry, as well as specific functional moieties and reactivity, are described in Thomas Sorrell, *Organic Chemistry*, University Science Books (Sausalito, CA), 1999; and Kemp and Vellaccio, *Organic Chemistry*, Worth Publishers, Inc. (New York), 1980; the entire contents of which are incorporated herein by reference.

**[0014]** The term “aliphatic”, as used herein, includes both saturated and unsaturated, straight chain (*i.e.*, unbranched), branched, acyclic, cyclic, or polycyclic aliphatic hydrocarbons, which are optionally substituted with one or more functional

groups. As will be appreciated by one of ordinary skill in the art, "aliphatic" is intended herein to include, but is not limited to, alkyl, alkenyl, alkynyl, cycloalkyl, cycloalkenyl, and cycloalkynyl moieties. Thus, as used herein, the term "alkyl" includes straight, branched and cyclic alkyl groups. An analogous convention applies to other generic terms such as "alkenyl", "alkynyl", and the like. Furthermore, as used herein, the terms "alkyl", "alkenyl", "alkynyl", and the like encompass both substituted and unsubstituted groups. In certain embodiments, as used herein, "lower alkyl" is used to indicate those alkyl groups (cyclic, acyclic, substituted, unsubstituted, branched or unbranched) having 1-6 carbon atoms.

**[0015]** In certain embodiments, the alkyl, alkenyl, and alkynyl groups employed in the invention contain 1-20 aliphatic carbon atoms. In certain other embodiments, the alkyl, alkenyl, and alkynyl groups employed in the invention contain 1-10 aliphatic carbon atoms. In yet other embodiments, the alkyl, alkenyl, and alkynyl groups employed in the invention contain 1-8 aliphatic carbon atoms. In still other embodiments, the alkyl, alkenyl, and alkynyl groups employed in the invention contain 1-6 aliphatic carbon atoms. In yet other embodiments, the alkyl, alkenyl, and alkynyl groups employed in the invention contain 1-4 carbon atoms. Illustrative aliphatic groups thus include, but are not limited to, for example, methyl, ethyl, n-propyl, isopropyl, cyclopropyl, -CH<sub>2</sub>-cyclopropyl, vinyl, allyl, n-butyl, sec-butyl, isobutyl, tert-butyl, cyclobutyl, -CH<sub>2</sub>-cyclobutyl, n-pentyl, sec-pentyl, isopentyl, tert-pentyl, cyclopentyl, -CH<sub>2</sub>-cyclopentyl, n-hexyl, sec-hexyl, cyclohexyl, -CH<sub>2</sub>-cyclohexyl moieties and the like, which again, may bear one or more substituents. Alkenyl groups include, but are not limited to, for example, ethenyl, propenyl, butenyl, 1-methyl-2-buten-1-yl, and the like. Representative alkynyl groups include, but are not limited to, ethynyl, 2-propynyl (propargyl), 1-propynyl, and the like.

**[0016]** In general, the terms "aryl" and "heteroaryl", as used herein, refer to stable mono- or polycyclic, heterocyclic, polycyclic, and polyheterocyclic unsaturated moieties having preferably 3-14 carbon atoms, each of which may be substituted or unsubstituted. Substituents include, but are not limited to, any of the previously mentioned substituents, *i.e.*, the substituents recited for aliphatic moieties, or for other moieties as disclosed herein, resulting in the formation of a stable compound. In



certain embodiments of the present invention, “aryl” refers to a mono- or bicyclic carbocyclic ring system having one or two aromatic rings including, but not limited to, phenyl, naphthyl, tetrahydronaphthyl, indanyl, indenyl, and the like. In certain embodiments of the present invention, the term “heteroaryl”, as used herein, refers to a cyclic aromatic radical having from five to ten ring atoms of which one ring atom is selected from S, O, and N; zero, one, or two ring atoms are additional heteroatoms independently selected from S, O, and N; and the remaining ring atoms are carbon, the radical being joined to the rest of the molecule via any of the ring atoms, such as, for example, pyridyl, pyrazinyl, pyrimidinyl, pyrrolyl, pyrazolyl, imidazolyl, thiazolyl, oxazolyl, isooxazolyl, thiadiazolyl, oxadiazolyl, thiophenyl, furanyl, quinolinyl, isoquinolinyl, and the like.

**[0017]** It will be appreciated that aryl and heteroaryl groups can be unsubstituted or substituted, wherein substitution includes replacement of one, two, three, or more of the hydrogen atoms thereon independently with any one or more of the following moieties including, but not limited to: aliphatic; heteroaliphatic; aryl; heteroaryl; arylalkyl; heteroarylalkyl; alkoxy; aryloxy; heteroalkoxy; heteroaryloxy; alkylthio; arylthio; heteroalkylthio; heteroarylthio; -F; -Cl; -Br; -I; -OH; -NO<sub>2</sub>; -CN; -CF<sub>3</sub>; -CH<sub>2</sub>CF<sub>3</sub>; -CHCl<sub>2</sub>; -CH<sub>2</sub>OH; -CH<sub>2</sub>CH<sub>2</sub>OH; -CH<sub>2</sub>NH<sub>2</sub>; -CH<sub>2</sub>SO<sub>2</sub>CH<sub>3</sub>; -C(O)R<sub>x</sub>; -CO<sub>2</sub>(R<sub>x</sub>); -CON(R<sub>x</sub>)<sub>2</sub>; -OC(O)R<sub>x</sub>; -OCO<sub>2</sub>R<sub>x</sub>; -OCON(R<sub>x</sub>)<sub>2</sub>; -N(R<sub>x</sub>)<sub>2</sub>; -S(O)<sub>2</sub>R<sub>x</sub>; -NR<sub>x</sub>(CO)R<sub>x</sub>, wherein each occurrence of R<sub>x</sub> independently includes, but is not limited to, aliphatic, heteroaliphatic, aryl, heteroaryl, arylalkyl, or heteroarylalkyl, wherein any of the aliphatic, heteroaliphatic, arylalkyl, or heteroarylalkyl substituents described above and herein may be substituted or unsubstituted, branched or unbranched, cyclic or acyclic, and wherein any of the aryl or heteroaryl substituents described above and herein may be substituted or unsubstituted. Additional examples of generally applicable substituents are illustrated by the specific embodiments shown in the Examples that are described herein.

**[0018]** The term “cycloalkyl”, as used herein, refers specifically to groups having three to seven, preferably three to ten carbon atoms. Suitable cycloalkyls include, but are not limited to cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloheptyl and the like, which, as in the case of other aliphatic, heteroaliphatic, or heterocyclic moieties, may optionally be substituted with substituents including, but

not limited to aliphatic; heteroaliphatic; aryl; heteroaryl; arylalkyl; heteroarylalkyl; alkoxy; aryloxy; heteroalkoxy; heteroaryloxy; alkylthio; arylthio; heteroalkylthio; heteroarylthio; -F; -Cl; -Br; -I; -OH; -NO<sub>2</sub>; -CN; -CF<sub>3</sub>; -CH<sub>2</sub>CF<sub>3</sub>; -CHCl<sub>2</sub>; -CH<sub>2</sub>OH; -CH<sub>2</sub>CH<sub>2</sub>OH; -CH<sub>2</sub>NH<sub>2</sub>; -CH<sub>2</sub>SO<sub>2</sub>CH<sub>3</sub>; -C(O)R<sub>x</sub>; -CO<sub>2</sub>(R<sub>x</sub>); -CON(R<sub>x</sub>)<sub>2</sub>; -OC(O)R<sub>x</sub>; -OCO<sub>2</sub>R<sub>x</sub>; -OCON(R<sub>x</sub>)<sub>2</sub>; -N(R<sub>x</sub>)<sub>2</sub>; -S(O)<sub>2</sub>R<sub>x</sub>; -NR<sub>x</sub>(CO)R<sub>x</sub>, wherein each occurrence of R<sub>x</sub> independently includes, but is not limited to, aliphatic, heteroaliphatic, aryl, heteroaryl, arylalkyl, or heteroarylalkyl, wherein any of the aliphatic, heteroaliphatic, arylalkyl, or heteroarylalkyl substituents described above and herein may be substituted or unsubstituted, branched or unbranched, cyclic or acyclic, and wherein any of the aryl or heteroaryl substituents described above and herein may be substituted or unsubstituted. Additional examples of generally applicable substitutents are illustrated by the specific embodiments shown in the Examples that are described herein.

**[0019]** The term “heteroaliphatic”, as used herein, refers to aliphatic moieties that contain one or more oxygen, sulfur, nitrogen, phosphorus, or silicon atoms, *e.g.*, in place of carbon atoms. Heteroaliphatic moieties may be branched, unbranched, cyclic or acyclic and include saturated and unsaturated heterocycles such as morpholino, pyrrolidinyl, *etc.* In certain embodiments, heteroaliphatic moieties are substituted by independent replacement of one or more of the hydrogen atoms thereon with one or more moieties including, but not limited to aliphatic; heteroaliphatic; aryl; heteroaryl; arylalkyl; heteroarylalkyl; alkoxy; aryloxy; heteroalkoxy; heteroaryloxy; alkylthio; arylthio; heteroalkylthio; heteroarylthio; -F; -Cl; -Br; -I; -OH; -NO<sub>2</sub>; -CN; -CF<sub>3</sub>; -CH<sub>2</sub>CF<sub>3</sub>; -CHCl<sub>2</sub>; -CH<sub>2</sub>OH; -CH<sub>2</sub>CH<sub>2</sub>OH; -CH<sub>2</sub>NH<sub>2</sub>; -CH<sub>2</sub>SO<sub>2</sub>CH<sub>3</sub>; -C(O)R<sub>x</sub>; -CO<sub>2</sub>(R<sub>x</sub>); -CON(R<sub>x</sub>)<sub>2</sub>; -OC(O)R<sub>x</sub>; -OCO<sub>2</sub>R<sub>x</sub>; -OCON(R<sub>x</sub>)<sub>2</sub>; -N(R<sub>x</sub>)<sub>2</sub>; -S(O)<sub>2</sub>R<sub>x</sub>; -NR<sub>x</sub>(CO)R<sub>x</sub>, wherein each occurrence of R<sub>x</sub> independently includes, but is not limited to, aliphatic, heteroaliphatic, aryl, heteroaryl, arylalkyl, or heteroarylalkyl, wherein any of the aliphatic, heteroaliphatic, arylalkyl, or heteroarylalkyl substituents described above and herein may be substituted or unsubstituted, branched or unbranched, cyclic or acyclic, and wherein any of the aryl or heteroaryl substituents described above and herein may be substituted or unsubstituted. Additional examples of generally applicable substitutents are illustrated by the specific embodiments shown in the Examples that are described herein.

**[0020]** The term, *associated with*, is used to describe the interaction between or among two or more groups, moieties, compounds, monomers, *etc.* When two or more entities are “associated with” one another as described herein, they are linked by a direct or indirect covalent or non-covalent interaction. Preferably, the association is covalent. The covalent association may be through an amide, ester, carbon-carbon, disulfide, carbamate, ether, or carbonate linkage. The covalent association may also include a linker moiety such as a cleavable linker (*e.g.*, disulfide linker, photocleavable linker, *etc.*). Desirable non-covalent interactions include hydrogen bonding, van der Waals interactions, hydrophobic interactions, magnetic interactions, electrostatic interactions, *etc.*

**[0021]** *Polynucleotide*, *nucleic acid*, or *oligonucleotide* refers to a polymer of nucleotides. The polymer may include natural nucleosides (*i.e.*, adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine), nucleoside analogs (*e.g.*, 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine), chemically modified bases, biologically modified bases (*e.g.*, methylated bases), intercalated bases, modified sugars (*e.g.*, 2'-fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose), or modified phosphate groups (*e.g.*, phosphorothioates and 5'-N-phosphoramidite linkages).

**[0022]** A *protein* comprises a polymer of amino acid residues linked together by peptide bonds. The term, as used herein, refers to proteins, polypeptides, and peptide of any size, structure, or function. Typically, a protein will be at least three amino acids long. A protein may refer to an individual protein or a collection of proteins. A protein may refer to a full-length protein or a fragment of a protein. Inventive proteins preferably contain only natural amino acids, although non-natural amino acids (*i.e.*, compounds that do not occur in nature but that can be incorporated into a polypeptide chain) and/or amino acid analogs as are known in the art may alternatively be employed. Also, one or more of the amino acids in an inventive protein may be modified, for example, by the addition of a chemical entity such as a

carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, *etc.* A protein may also be a single molecule or may be a multi-molecular complex. A protein may be just a fragment of a naturally occurring protein or peptide. A protein may be naturally occurring, recombinant, or synthetic, or any combination of these.

[0023] The term *small molecule*, as used herein, refers to a non-peptidic, non-oligomeric organic compound either synthesized in the laboratory or found in nature. Small molecules, as used herein, can refer to compounds that are “natural product-like”, however, the term “small molecule” is not limited to “natural product-like” compounds. Rather, a small molecule is typically characterized in that it possesses one or more of the following characteristics including having several carbon-carbon bonds, having multiple stereocenters, having multiple functional groups, having at least two different types of functional groups, and having a molecular weight of less than 1500, although this characterization is not intended to be limiting for the purposes of the present invention.

### Brief Description of the Drawing

[0024] *Figure 1* is a comparison of the substrate organization used in a two-pool reaction discovery system versus a single-pool reaction discovery system.

[0025] *Figure 2* shows a selection for bond forming reaction in the inventive single-pool reaction discovery system.

[0026] *Figure 3* shows the analysis of polynucleotide sequences encoding bond-forming reactions.

[0027] *Figure 4* illustrates an exemplary method of assembling the template molecules with attached substrates in the inventive reaction discovery system.

[0028] *Figure 5* shows the results of a model selection in the presence of  $\text{Na}_2\text{PdCl}_4$ .

[0029] *Figure 6* shows the design of an experiment to test a new architecture with a shorter template, with a linker of 6 carbon atoms and a distance between the two substrates of 5 bases.

[0030] *Figure 7* shows exemplary substrate structures.

- [0031] *Figure 8* shows details of a matrix assembly strategy.
- [0032] *Figure 9* shows component preparation for matrix assembly.
- [0033] *Figure 10* shows B-oligos labeled with pool B substrates B1-B14.
- [0034] *Figure 11* shows B-oligos labeled with pool A substrates A1-A14.
- [0035] *Figure 12* shows A-oligos labeled with pool A substrates.
- [0036] *Figure 13* shows A-oligos labeled with pool B substrates.
- [0037] *Figure 14* shows an exemplary assembly of templates with two substrates attached useful in the inventive reaction discovery system. The assembly includes adding the  $A_n$  coding region onto an oligonucleotide with the substrate  $B_m$  attached using a primer extension reaction. The substrate  $A_n$  with its biotin disulfide linker is then ligated onto the 5' end to form the full template.
- [0038] *Figure 15* demonstrates the use of microarray technology to analyze the results of the bonding forming reaction conditions. Each spot on the DNA microarray corresponds to an  $A_m \times B_n$  encoding oligo template. Arrays were printed using a Genemachines OmniGrid instrument in the Bauer Center for Genomics Research at Harvard University.
- [0039] *Figure 16* shows a microarray analysis used to validate the inventive reaction discovery system by screening for known bond forming reactions.
- [0040] *Figure 17* demonstrates the use of the reaction discovery system to screen for bond forming reaction mediated by Cu(I). Reaction conditions included 1 mM Cu(I) for 10 minutes at 25 °C in four different aqueous solvent mixtures: 90% acetonitrile ( $\text{CH}_3\text{CN}$ ), 80% DMF, 90% methanol, and 90% dioxane.
- [0041] *Figure 18* demonstrates further use of the inventive reaction discovery system to explore enamine chemistry. The pyrrolidine in the reactions is expected to mediate an enamine aldol reaction (*e.g.*, between ketone (A12) and aryl aldehyde (B2).
- [0042] *Figure 19* demonstrates further use of the inventive reaction discovery system to explore reductive amination chemistry.
- [0043] *Figure 20* demonstrates the use of the reaction discovery system to screen for bond forming reaction mediated by Pd(II). Reaction conditions included 1 mM Pd(II) in MOPS buffer pH 7 for 20 minutes at four different temperatures: 25 °C, 37 °C, 50 °C, and 65 °C.

[0044] *Figure 21* demonstrates the use of the reaction discovery system to screen for bond forming reaction mediated by Pd(II). Reaction conditions included 1 mM Pd(II) in 90% acetonitrile for 20 minutes at four different temperatures: 37 °C, 50 °C, 65 °C, and 95 °C.

[0045] *Figure 22* demonstrates the use of the reaction discovery system to screen for bond forming reaction mediated by Pd(II). Reaction conditions included 1 mM Pd(II) in 90% DMSO for 20 minutes at four different temperatures: 37 °C, 50 °C, 65 °C, and 95 °C.

### Detailed Description of Certain Preferred Embodiments of the Invention

[0046] The present invention provides a system for discovering new chemical reactions. This novel system for discovering new chemical reactivity and reactions is not encumbered by conventional wisdom with regard to functional group reactivity and allows for the examination of a broad range of both reaction conditions and substrates in a highly efficient manner. The inventive method of discovering new chemical reactions and chemical reactivity has several advantages over existing methods. For example, several groups have developed high-throughput screens to test the efficiency of a particular reaction under a variety of conditions (Kuntz *et al. Current Opinion in Chemical Biology* 3:313-319, 1999; Francis *et al. Curr. Opin. Chem. Biol.* 2:422-428, 1998; Pawlas *et al. J. Am. Chem. Soc.* 124:3669-3679, 2002; Lober *et al. J. Am. Chem. Soc.* 123:4366-4367, 2001; Evans *et al. Curr. Opin. Chem. Biol.* 6:333-338, 2002; Taylor *et al. Science* 280:267-270, 1998; Stambuli *et al. J. Am. Chem. Soc.* 123:2677-2678, 2001; each of which is incorporated herein by reference); however, these screens are limited to a small set of reaction types. Reactions have been analyzed in a high-throughput manner using fluorescence spectroscopy, colorimetric assay, thermographic analysis, and traditional chromatography (Dahmen *et al. Synthesis-Stuttgart* 1431-1449, 2001; Wennemers *Combinatorial Chemistry & High Throughput Screening* 4:273-285, 2001; each of which is incorporated herein by reference). Most high-throughput screens for chemical reactivity are useful for only a small set of reaction types because the screen depends on a particular property of the reaction such as the disappearance of an amine or the production of protons. As a result, high throughput screening methods can be useful for discovering catalysts for a

known or anticipated reason, but are poorly suited to discover novel reactivity different from a reaction of interest.

[0047] A non-biased search for chemical reactions would examine a broad range of both reaction conditions and substrates in a highly efficient manner that is practical on the scale of thousands of different reactions. In certain embodiments, the inventive system only requires nanomolar, picomolar, or femtomolar quantities of material per reaction discovery experiment. The inventive system for discovering novel chemical reactions offers a much greater chance of discovering unexpected and unprecedented reactivity that may lead to new insights into reactivity and to useful new reactions for chemical synthesis.

[0048] The inventive system also differs from a previously disclosed reaction discovery system (*see* U.S. patent application, U.S.S.N. 60/404,395, filed August 19, 2002; U.S.S.N. 10/643,752, filed August 19, 2003; U.S.S.N. 60/602,255, filed August 17, 2004; and U.S.S.N. 11/205,492, filed August 17, 2005; each of which is incorporated herein by reference) in that the instant system does not rely on nucleic acid hybridization to bring the two potential substrates together. Hybridization unfortunately limits the reaction conditions that can be explored using the previous system because duplex formation typically requires an aqueous solution with a relatively high salt concentration at a relatively low temperature.

#### *Non-DNA Templated Reaction Discovery*

[0049] The new system allows for a broader, non-biased search for chemical reactivity of a large number of diverse reactants in parallel. Although some chemical reactions are compatible with water as the solvent, the vast majority of ligands, catalysts, and reagents that are used by organic chemists are not compatible with an aqueous solvent and/or are not soluble in aqueous solutions. To access these reaction conditions, a new selection-based approach to reaction discovery was developed in which organic solvents or organic solvent/water mixtures can be used as solvent and higher temperature can be utilized. In addition, the new system requires less material per reaction discovery experiment than the DNA-templated approach. The new system, therefore, allows for a broader exploration of reactivity space than any previous systems for reaction discovery.

**[0050]** This new approach involves attaching substrate pairs on the same nucleic acid strand. The substrates are attached to the strand in such a manner that they are able to react with each other under suitable conditions and successful reaction of the substrate pair allows for selection and identification of the substrates. The reactivity of these substrates under many different conditions (*e.g.*, solvent, catalysts, pH, reagents, *etc.*) may be evaluated. The sequence of the template molecule is used to identify the substrates attached to the template. The present invention also includes any new chemical reactions that are discovered using the inventive method and system of discovering new chemical reactions.

**[0051]** The inventive system first involves preparing a template with a substrate pair attached. The template is then subjected to test reaction conditions, and the template is subsequently selected if the reaction conditions have effected a direct covalent attachment between the two substrates on the template. The sequence of the template is then optionally determined to identify the successfully reacting substrates. In certain embodiments, DNA microarray technology is used to identify the reacted substrate pairs. The system is particularly amenable for analyzing many different combinations of substrates and reaction conditions in parallel. The system also allows for the testing of a library of different template molecules to be performed in one pot.

**[0052]** The template typically includes two substrates and sequences that identify the attached substrates. The template may be any nucleic acid molecule. In certain embodiments, the template is DNA. In other embodiments, the template is RNA. In other embodiments, the template is a derivative of DNA or RNA. For example, the template may include unnatural bases. In certain embodiments, an unnatural bases is used to attach one of the substrates to the template. The two substrates may be attached anywhere along the template molecule. The substrates are typically attached in such a manner that they will react under suitable conditions. In certain embodiments, at least one substrate is attached at an end of the template. In other embodiments, at least one substrate it attached to an internal base of the template. For example, one of the substrates may be attached to a modified base such as amino-modified deoxythymidine. In certain particular embodiments, one substrate is attached to an end of the template, and the other substrate is attached to an internal base of the template. In other embodiments, both substrates are attached to an internal



base of the template. In still other embodiments, both substrates are attached to the same or opposite ends of the template. The two substrates are usually attached at less than 100, 75, 50, 40, 30, 25, 20, 15, 10, or 5 bases from each other. In certain particular embodiments, the two substrates are attached at a distance between 1 and 20 bases from each other. In certain particular embodiments, the two substrates are attached at a distance between 5 and 15 bases from each other. In certain embodiments, the substrate is attached to the template through a linker. In certain embodiments, the linker contains 1-20 carbons or heteroatoms. In certain particular embodiments, the linker contains 1-10 carbons or heteroatoms. In certain embodiments, the linker contains approximately 6 carbon atoms or heteroatoms. In certain embodiments, the linker is substituted. In other embodiments, the linker is unsubstituted. In certain embodiments, the linker include cyclic structures. In other embodiments, the linker does not include cyclic structures. In certain embodiments, the linker is cleavable. Exemplary cleavable linkers include disulfide bonds, ester bonds, amide bonds, *etc.* In certain embodiments, a disulfide bond is used to link the substrate to the end of the template. The linker may also include an affinity agent such as biotin. Other affinity agents useful in the present invention include polyhistidine, antibody, fragments of antibodies, epitopes, *etc.* In certain embodiments, when the linker is cleaved the substrate attached through the linker to the template and the affinity agent are released from the template as shown in *Figures 2-3*.

**[0053]** The substrates attached to the template may include any chemical functional group. Particularly interesting are reactive functional groups that have been shown to be useful in other chemical reactions. Functional groups that have been shown to be useful in carbon-carbon bond forming reactions are particularly useful. Exemplary substrates include aliphatic halides (*e.g.*, alkyl halides, akenyl halides, akynyl halides), aryl halides, esters, amides, carbonates, carbamates, ureas, alcohols, thiols, amines (*e.g.*, aliphatic amines, aryl amines, dialiphatic amines, trialiphatic amines, *etc.*), alkyls, alkenes, alkynes, aryls, heteroaryls, phosphorus-containing groups (*e.g.*, phosphonium salts), *etc.*

**[0054]** The template molecule typically contains sequence that encode the identity of the substrates attached to the template (*i.e.*, encoding sequences).

Depending on the number of substrates, the identity of the attached substrates is encoded in 2, 3, 4, 5, 6, 7, 8, 9, 10, or more bases. Each substrate used in a library is associated with its own identifying sequences. The template also may contain other sequences useful in the present invention. For example, the template may include primer sequences. In certain embodiments, the primer sequences are useful for amplifying the template nucleic acid by the PCR. In other embodiments, the primer sequences are useful for determining the sequence of the template. The template may also contain linking sequences that link the various encoding sequences or primer sequences together.

**[0055]** Once the template is prepared with its substrate combination, it is exposed to a particular set of reaction conditions. In certain embodiments, the template is exposed to one set of reaction conditions. The template may also be exposed to a sequence of multiple reaction conditions. The reaction conditions may include solvent, pH, catalyst, ligand, salt concentration, stoichiometric reagent, activating reagent, deprotecting reagent, protecting reagent, temperature, pressure, duration of reaction, presence of water, presence of oxygen, presence of another gas, presence of a metal, presence of a surface, presence of an ion, *etc.* In certain embodiments, the reaction conditions include an organometallic catalyst (*e.g.*, Pd, Pt, Co, Mo, Cu, Zn, Os, Hg, *etc.*). In certain particular embodiments, the reaction conditions include a catalyst that has been shown to be useful in carbon-carbon bond forming reactions. In other embodiments, the reaction conditions include the addition of an acid or base. In certain embodiments, the reaction conditions include an acidic pH (<7). In other embodiments, the reaction conditions include a basic pH (>7). In still other embodiments, the reaction conditions include a neutral pH (approximately 7). In certain embodiments, the reaction conditions include a chiral reagent.

**[0056]** In certain embodiments, the reaction conditions include an organic solvent. Exemplary organic solvent useful in the present invention include acetonitrile, tetrahydrofuran (THF), chloroform, methylene chloride, dimethylformamide (DMF), DMSO, dioxane, benzene, toluene, diethyl ether, hexanes, ethanol, methanol, *etc.* The solvent used may also include a mixture of an organic solvent and water. In certain embodiments, the reaction conditions include a mixed aqueous-organic solvent including acetonitrile, DMF, DMSO, methanol, or

dioxane. The percentage of water in the mixture may range from 0% to 50%. In certain embodiments, the percentage of water in the mixture is approximately 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, or 50%. In certain embodiments, the reaction mixture include no water or at least as little water as is reasonable possible. In other embodiments, the reaction conditions include an aqueous system. In certain particular embodiments, the aqueous system is buffered at a particular pH.

[0057] In certain embodiments, the reactions conditions include a particular temperature. The inventive system is particular useful to explore reaction condition at higher than ambient temperature because the inventive system does not require duplex formation. In certain embodiments, temperature ranges from -78 °C to 200 °C. In other embodiments, the temperature ranges from 0 °C to 200 °C. In yet other embodiments, the temperature ranges from 25 °C to 200 °C. In yet other embodiments, the temperature ranges from 30 °C to 200 °C. In yet other embodiments, the temperature ranges from 20 °C to 100 °C. In yet other embodiments, the temperature ranges from 25 °C to 100 °C. In yet other embodiments, the temperature ranges from 30 °C to 100 °C. In certain embodiments, the temperature of the reaction conditions is approximately 10 °C, 20 °C, 30 °C, 40 °C, 50 °C, 60 °C, 70 °C, 80 °C, 90 °C, 100 °C, 110 °C, 120 °C, 130 °C, 140 °C, 150 °C, 160 °C, 170 °C, 180 °C, 190 °C, or 200 °C. In certain embodiments, the temperature of the reaction conditions is approximately 25 °C. In certain embodiments, the temperature of the reaction conditions is approximately 37 °C. In certain embodiments, the temperature of the reaction conditions is approximately 50 °C. In certain embodiments, the temperature of the reaction conditions is approximately 65 °C. In certain embodiments, the temperature of the reaction conditions is approximately 95 °C. In certain embodiments, the temperature of the reaction conditions is less than 100 °C.

[0058] After the template has been exposed to the desired test reaction conditions, the cleavable linker is cleaved using an appropriate reagent. If the reaction conditions have not resulted in the formation of a direct covalent linkage between the two substrates, cleavage of the linker will result in one of the substrates and the affinity reagent being completely removed from the template. In the case of a

disulfide linkage being used, any reducing reagent may be used to cleave the linkage. In certain embodiments, high concentrations of tris-(2-carboxyethyl)phosphine hydrochloride are used to cleave the disulfide linker. For an ester or amide linker, an esterase or amidase may be used to cleave the linker, respectively. In the case of an amide linker, a protease may also be used to cleave the linker. Such ester and amide linkers may also be cleaved by acid or base hydrolysis. The chemistry used to cleave the linker should not modify or cleave any covalent bond that may have formed between the two substrates. Those templates that include substrate combinations that have reacted to form a covalent bond will remain covalently linked to the affinity reagent through one of the substrates (*e.g.*, attached through the internal modified base). The templates with substrate combinations that have reacted can then be isolated using a resin known to bind the affinity reagent. For example, with a biotin affinity reagent, streptavidin beads can be used to isolate the templates with substrate combinations that have reacted to form a covalent bond.

**[0059]** Once the templates with substrate combinations that have reacted are isolated, the attached nucleic acid sequence may be analyzed to determine the identity of the substrates. In certain embodiments, the template nucleic acid is first amplified by the PCR. The sequence of the nucleic acid is then determined by traditional sequencing of the nucleic acid, by micro-array based methods, by mass spectral analysis, or by any other methods used to determine the sequence of a nucleic acid. The PCR amplification allows a researcher to perform selection on pmol quantities of material so that even a single nmol-scale preparation of the original template can provide enough material for the testing of thousands of different reactions conditions.

**[0060]** In other embodiments, microarray technology is used to identify the substrate pairs that results in bond formation under the test reaction conditions. The analysis using microarray technology is illustrated in *Figure 15-22*. DNA microarray are printed with oligonucleotide probes that correspond to each encoding oligonucleotide in a library. For example, the probes may correspond to each permutation of substrates including homocoupling. In certain embodiments, the microarray also includes standards. In certain embodiments, the array includes positive controls that correspond to known reaction. In certain embodiments, the array includes negative controls that correspond to substrates that are known not to

react under the test reaction conditions. The selected oligonucleotides with substrate pairs that have reacted are then allowed to hybridize to the array, and spots with hybridized probe indicate that a reaction took place between the substrates under the test reaction conditions.

**[0061]** The inventive reaction discovery system combines *in vitro* selection with the use of nucleic acid technology to efficiently search for novel bond-forming reaction independent of reactant structures. The ability to select directly for covalent bond formation, the minute scale required for analysis (*e.g.*, 1 pmol total material of template molecules per discovery experiment), and compatibility of the system with a wide variety of reaction conditions enables the search for unprecedented reactivity that can examine thousands of combinations of reactants and reaction conditions in one or several experiments as shown in *Figures 17-22*.

**[0062]** In certain embodiments, a library of template molecules with various combinations of substrates are prepared. See, for example, *Figure 7*. In certain embodiments, all possible combinations of a set of substrates are provided for in a library of template molecules. The library of templates may contain at least 10 members, 20 members, 30 members, 50 members, 100 members, 250 members, 500 members, or 1,000 members. In certain embodiments, all the members of a library are combined in one pot and subjected to a specific set of reaction conditions. In certain embodiments, 0.1-100 pmol of total material of library templates are used in a round of reaction discovery. In certain embodiments, 1-50 pmol of material are used. In other embodiments, 1-10 pmol are used. In yet other embodiments, approximately 1 pmol of material is used in a reaction discovery experiment. Each of the members of the library may be subjected to tens, hundreds, or thousands of different reaction conditions.

#### *Preparation of Template Molecules*

**[0063]** The present invention also provides methods of preparing the templates useful in the present reaction discovery system. Any method may be used to prepare the templates with the pair of substrates, the linker, affinity reagent, and identifying sequences. Techniques are known in the art for preparing DNA molecules of any sequence and modifying those sequences (Ausubel *et al.*, eds., *Current*

*Protocols in Molecular Biology*, 1987; Sambrook *et al.* *Molecular Cloning: A Laboratory Manual*, 2nd Ed., 1989; each of which is incorporated herein by reference). For example, the DNA molecule may be prepared by a DNA synthesizer and subsequently modified at the termini or at an internal base. In certain embodiments, a modified internal base is inserted into the sequence to be later modified by attaching a substrate. In certain embodiments, two modified internal bases, which are later modified by attaching a substrate, are inserted into the sequence at a pre-determined distance apart (*e.g.*, 1-15 bases apart).

**[0064]** In certain embodiments, the method of preparing of the template is a modification of the method use to prepare templates in the two-pool reaction discovery system as described in U.S. application, U.S.S.N. 11/205,493, filed August 17, 2005, incorporated herein by reference. This methods generally involves labeling an oligonucleotide with a single substrate and using enzymatic steps to assemble the full-length template. As shown in *Figure 4*, the DNA with the internal substrate B<sub>1</sub> is prepared. This DNA is then annealed to another sequence encoding the A<sub>1</sub> substrate and by a primer extension reaction the encoding sequence for substrate A<sub>1</sub> is added to the sequence encoding the substrate B<sub>1</sub>. The resulting DNA template contains the coding region for both substrates and is linked to one substrate of the pair. In the next step, a simple ligation appends an oligonucleotide with substrate A<sub>1</sub> attached to its 5' end via a cleavable affinity labeled linker to the 5' end of the template from the first step. The resulting full-length template contains the two coding regions for the two substrates. One substrate A<sub>1</sub> is attached to the 5' end through a cleavable linker with an affinity tag, and the other substrate B<sub>1</sub> is attached at an internal modified base. The full-length template with its combination of substrates may then be optionally purified using any technique known in the art including denaturing PAGE, HPLC, or column chromatography. In certain embodiments, the method is amenable to preparing a library of individual templates with substrates A<sub>1</sub>-A<sub>n</sub> and substrates B<sub>1</sub>-B<sub>n</sub>. The template molecules are prepared individually or in parallel. In certain embodiments, the template molecules are prepared in parallel with 10-100 different template molecules being prepared at once. In certain embodiments, the template molecules are prepared in parallel with 10-30 different template molecules being prepared at once.

[0065] The resulting template molecules are also considered an aspect of the invention. In certain embodiments, the template molecule comprises an oligonucleotide with sequences to identify the attached substrates, two substrates, a cleavable linker attaching at least one of the substrates to the oligonucleotide, and an affinity agent for use in the selection process. In certain embodiments, the oligonucleotide also include primer sequences useful in preparing the template, useful in PCR amplification, and/or useful in sequencing or microarray analysis.

#### *Kits*

[0066] The invention also provides kits for practicing the inventive system. In certain embodiments, the kit includes all the materials needed by a researcher to conduct a round of reaction discovery. The kit may include all or some of the following: oligonucleotides, nucleotides, modified nucleotides, linkers, affinity reagents (*e.g.*, biotin, antigens, epitopes, peptides, *etc.*), substrates, buffers, enzymes, materials for PAGE, columns, solvents, catalysts, reagents, ligands, microarrays, vials, Eppendorf tubes, instructions, *etc.* In certain embodiments, the kit may allow the user to provide his or her own substrates. In certain embodiments, the kit may allow the user to test his or her own reaction conditions. In certain embodiments, the materials in the kit are packaged conveniently for the researcher to use.

[0067] These and other aspects of the present invention will be further appreciated upon consideration of the following Examples, which are intended to illustrate certain particular embodiments of the invention but are not intended to limit its scope, as defined by the claims.

### **Examples**

#### **Example 1-Reaction Discovery in 100% Aqueous Solutions and Organic Solvent-Water Mixtures**

[0068] We tested the ability to select for reactions under conditions that do not favor duplex formation in a model experiment that included one sequence attached to

a reactive combination of substrates and another sequence attached to a non-reactive combination (*Figure 5*). The experiment was designed to mimic a reaction discovery selection wherein one substrate combination reacts to form a bond between the two substrates and the other combinations do not react. We separately prepared a template linked to a terminal alkene and a terminal alkyne (template T1) and a template linked to two alkanes (template T2) using the method described above. Fifteen bases separate the 5' end of the template to which one substrate is attached from the modified dT to which the other substrate is attached, a distance comparable to the separation between the 5' end of the template and the 3' end of the complementary DNA in the two-pool system. T1 contains a restriction site for the endonuclease *Ava I* that is not present in T2. T1 was combined with a 100-fold excess of T2 in an aqueous NaCl solution or a water-acetonitrile mixture. The solutions were treated with 500  $\mu\text{M}$   $\text{Na}_2\text{PdCl}_4$ , conditions under which an alkene and an alkyne react to form an enone, and selections for bond formation were performed as described above. The selected sequences were amplified by PCR and digested with *Ava I* to determine the enrichment of the sequence encoding the alkene and alkyne. Selection for bond formation in aqueous NaCl provided a  $\sim 130$ -fold enrichment of T1. Importantly,  $> 50$ -fold enrichment of T1 was observed when selections were performed in either 50% ACN- $\text{H}_2\text{O}$  or in 90% ACN- $\text{H}_2\text{O}$  with 100  $\mu\text{M}$  cetyl trimethylammonium bromide to enhance DNA solubility (Ijiro *et al.* A DNA-lipid complex soluble in organic solvents. *J. Chem. Soc. Chem. Comm.* **18**, 1339-1341 (1992); Tanaka *et al.* A DNA-lipid complex in organic media and formation of an aligned cast film. *J. Am. Chem. Soc.* **118**, 10679-10683 (1996); each of which is incorporated herein by reference). A control selection in which  $\text{Na}_2\text{PdCl}_4$  was omitted showed no enrichment of T1.

[0069] Since the selection is designed only to separate biotin-linked sequences from non-biotin-linked sequences, it is possible that the enrichment observed for T1 upon exposure to  $\text{Na}_2\text{PdCl}_4$  is a result of the alkene reacting with functionality on the DNA and not with the alkyne. To determine the extent to which this reactivity accounts for the enrichment of T1 in the experiment described above, we performed a selection where T1 was replaced with T3, a template with the same sequence as T1 but in which the alkyne was replaced with a ketone, a substrate that is not reactive



with the alkene under the reaction conditions (Chapter 3). Under identical selection conditions, the enrichment of T3 was much less than that observed for T1 (*Figure 5*). T3 was enriched < 20 fold in 0.1 M NaCl and < 10-fold in both 50% ACN/H<sub>2</sub>O and 90% ACN/H<sub>2</sub>O. These results demonstrate that the enrichment observed for T1 is primarily a result of a bond-formation between the alkyne and the alkene and to a minor extent a result of the alkene reacting with functionality on the DNA. We speculate that Wacker-type addition of nucleophilic exocyclic amines on the DNA to a Pd-activated alkene accounts for low levels of bond-formation between DNA and the alkene.

[0070] The enrichment factors for T1 are modest considering the high affinity of the biotin-streptavidin interaction. Enrichment may be negatively affected by incomplete cleavage of the disulfide bonds, a low reactivity of the two substrates separated by 15 bases, or a Pd-specific increase in the background. Since a microarray-based analysis of a reaction discovery selection evaluates reactivity based on the relative abundance of individual sequences before and after selection (Chapter 3), we anticipate that the enrichment factors observed for the single-pool system in these model selections will yield readily detectable signals in an array analysis. Decreasing the separation between the two substrates may also increase the signal arising from reactive substrate combinations.

## Experimental Methods

[0071] **General Methods.** DNA synthesis was carried out using standard reagents and protocols as described in Chapter 2 with exceptions noted below. Oligonucleotides were cleaved off of the CPG resin using AMA treatment for 10 min at 65° C, purified by reversed-phase HPLC and quantitated by UV spectroscopy. The A-labeled and B-labeled oligonucleotides were prepared using the indicated carboxylic acids and the labeling procedures described in Chapter 3. Labeled oligonucleotides were characterized by MALDI-TOF MS and observed masses were within 0.075% of calculated masses. The preparations of the A-coding oligonucleotide and splint oligonucleotides used in the template assemblies were described in Chapter 3.

**[0072] Preparation of A-Labeled Oligonucleotides.** DNA synthesis was carried out using standard reagents and monomers and three modified phosphoramidites: 5' Amino-Modifier-5, Biotin phosphoramidite, and Thiol-Modifier C6 S-S (all from Glen Research). The standard protocol was modified to include double deblocking and triple capping for the cycle incorporating the thiol-modifier phosphoramidite. This modification was made to minimize the possibility of truncation byproducts that lack the disulfide linkage and are therefore permanently linked to biotin. The oligonucleotide was labeled with hexanoic acid (Aldrich) (calculated mass: 4044.90; observed mass:  $4045 \pm 3$ ) and 6-hexenoic acid (Aldrich) (calculated mass: 4056.90; observed mass:  $4057 \pm 3$ ).

**[0073] Preparation of B-Labeled Oligonucleotides.** DNA synthesis was carried out using standard reagents and monomers and two modified phosphoramidites: Chemical Phosphorylating Reagent II and Amino-Modifier C6 dT (both from Glen Research). The sequence B1 contains a restriction site for *Ava I* and the sequence B2 lacks this site. The 5' phosphate group on each oligonucleotide was exposed with 2:1 H<sub>2</sub>O:concentrated NH<sub>4</sub>OH using the manufacturer's protocol (Glen Research) either before or after labeling. B1 was labeled with 6-heptynoic acid (Aldrich) (calculated mass: 9105.63; observed mass:  $9107 \pm 7$ ) and 6-oxoheptanoic acid (Aldrich) (calculated mass: 9123.64; observed mass:  $9123 \pm 7$ ). B2 was labeled with hexanoic acid (calculated mass: 9135.65; observed mass:  $9134 \pm 7$ ).

**[0074] Assembly of DNA Templates Linked to Two Substrates.** The full-length templates linked to two substrates were assembled in a two step sequence consisting of primer extension and ligation analogous to the modular assembly of pool A templates (*Figure 4*). Primer extension was typically performed on a 300 pmol scale (5  $\mu$ M A-coding oligonucleotide and 5  $\mu$ M B-labeled oligonucleotide) in 60  $\mu$ L at 25 °C for 1 h using Klenow *exo*<sup>-</sup> (New England BioLabs). Ligations were performed directly following buffer exchange of the primer extension reactions. Three-hundred pmol of A-labeled oligonucleotide and 300 pmol of the appropriate splint oligonucleotide were added to the buffer-exchanged reaction and the ligations were performed at 16° C for 2 h. Ligation reactions were ethanol precipitated and purified by denaturing PAGE according to standard procedures. Typically, a 300 pmol preparation yielded 100 pmol of purified full-length template. The following

templates were assembled with this procedure and used in the model selections: A-labeled 6-heptenoic acid + B1-labeled 6-heptynoic acid (T1), A-labeled hexanoic acid + B2-labeled hexanoic acid (T2), and A-labeled 6-heptenoic acid + B1-labeled 6-oxoheptanoic acid (T3).

**[0075] Selections with Na<sub>2</sub>PdCl<sub>4</sub>.** One pmol of T2 and 10 fmol of either T1 or T3 were combined in the solvents indicated in *Figure 5* and exposed to 500 μM Na<sub>2</sub>PdCl<sub>4</sub>. The total volume in each case was 30 μL. After 20 min at 25 °C, 10 μL of 0.1 M dithiothreitol (Aldrich) was added to the solution, followed by 150 μL of 0.1 M TCEP in 1.0 M phosphate pH 8.0. Disulfide cleavage was allowed to proceed for 1 h at 25 °C. The solution was diluted with 150 μL H<sub>2</sub>O and an aliquot of streptavidin magnetic particles (Roche Biosciences) with a 22 pmol binding capacity was added directly. Binding to streptavidin was allowed to proceed for 10 min and then the supernatant was removed and the particles were washed twice with 200 μL of 1.0 M NaCl, 10 mM Tris, 1 mM EDTA pH 7.5 ("wash buffer"). The particles were resuspended in 20 μL of 95% formamide- 5% 10 mM EDTA pH 8.0 ("elution solution") and heated at 65°C for 10 min. Fifteen μL of this eluant was added directly to 100 μL of fresh 0.1 M TCEP in 1.0 M phosphate pH 8.0. After 40 min at 25 °C, the solution was diluted with 100 μL H<sub>2</sub>O and a fresh aliquot of streptavidin particles was added directly. The supernatant was removed after 10 min and the particles were washed twice with 200 μL wash buffer. The particles were resuspended in 20 μL of elution solution and heated at 90° C for 10 min. Fifteen μL of this eluant was added to 45 μL of H<sub>2</sub>O and the resulting solution was passed through a gel filtration spin column (Princeton Separations) to remove formamide. The eluant was used directly in PCR reactions.

**[0076] PCR and *Ava I* Digest Analysis of Selections.** For each selection, a 2.5 μL aliquot of the final eluant (Selections with Na<sub>2</sub>PdCl<sub>4</sub>) was added to a 50 μL PCR reaction containing 2.5 mM MgCl<sub>2</sub>, 0.2 M dNTPs, and 500 nM of each primer. The sequences were amplified with 25 cycles of 95° C for 30 s, 55° C for 30 s, and 72° C for 20 s. Five μL aliquots of the PCR reactions were run on a 3% agarose gel to determine relative amounts of PCR products and the remaining 45 μL of each PCR reaction was precipitated with ethanol. Typically, one-fourth of the precipitated material (approximately 5 pmol of PCR product) was digested with 10 U of *Ava I*



**Example 2 – General Reaction Discovery Experimental**

**[0078]** The single pool reaction discovery starting material pool was assembled by combining 224 unique DNA template molecules at equal concentration and an internal standard at one tenth that concentration. This reaction discovery pool was prepared at a concentration of 0.5  $\mu$ M and stored at -20 °C. In a typical reaction discovery experiment, 1 pmol of total material (*i.e.*, 2  $\mu$ L of the starting reaction discovery pool) was combined with combined at the designated concentrations and solvents as noted in a 100  $\mu$ L, 200  $\mu$ L, or 300  $\mu$ L total volume. For example, in the case of the palladium (II) chemistry, 2  $\mu$ L of starting pool was combined with 10  $\mu$ L of Pd(II) (20 mM Na<sub>2</sub>PdCl<sub>4</sub> in double distilled water), 8  $\mu$ L of doubly distilled water, and 180  $\mu$ L of organic solvent (either acetonitrile or DMSO). Results for Pd(II) chemistry experiments are shown in *Figures 20-22*. In the case of the aqueous reaction reactions, the volume was adjusted with 100 mM MOPS buffer (pH 7) with 1 M NaCl in double distilled water. The reactions were then incubated at the designated temperature for the designated amount of time. After the reactions were complete, the DNA template molecules were precipitated with ethanol, and selections for bond formation were carried out. See *Figures 17-22*.

**Other Embodiments**

**[0079]** The foregoing has been a description of certain non-limiting preferred embodiments of the invention. Those of ordinary skill in the art will appreciate that various changes and modifications to this description may be made without departing from the spirit or scope of the present invention, as defined in the following claims.

## Claims

What is claimed is:

1. A method of identifying new chemical reactions, the method comprising steps of:  
providing one or more templates, which one or more templates have a first and second substrate associated therewith and the one or more templates including sequences that identify the substrates associated therewith;  
subjecting the templates with associated substrates to reaction conditions; and  
identifying a reaction product between the two substrates associated with the template.
2. The method of claim 1, wherein the template comprises a nucleic acid.
3. The method of claim 1, wherein the template comprises DNA or RNA.
4. The method of claim 1, wherein the template comprises DNA.
5. The method of claim 1, wherein one of the two substrates is attached to the end of the template.
6. The method of claim 1, wherein one of the two substrates is attached to the template through an internal nucleotide.
7. The method of claim 1, wherein the two substrates are within 20 bases of each other.
8. The method of claim 1, wherein the two substrates are within 15 bases of each other.
9. The method of claim 1, wherein the two substrates are within 10 bases of each

other.

10. The method of claim 1, wherein the substrates are covalently associated with the template.
11. The method of claim 1, wherein the methods does not require the hybridization of oligonucleotides.
12. The method of claim 1, wherein the step of subjecting the template comprises heating, cooling, adding a catalyst, adding a reagent, changing pH, increasing pressure, decreasing pressure, and adding a solvent.
13. The method of claim 1, wherein the step of subjecting the template comprises adding a stereoselective catalyst.
14. The method of claim 1, wherein the template further comprises an affinity reagent.
15. The method of claim 14, wherein the affinity reagent is biotin.
16. The method of claim 14, wherein the first substrate and the affinity reagent are attached to the template through a cleavable linker.
17. The method of claim 16, wherein the cleavable linker is a disulfide bond.
18. The method of claim 15, wherein the step of identifying comprises selecting templates covalently bound to biotin through the covalently attached substrates using streptavidin.
19. The method of claim 1, wherein the step of identifying comprises using DNA microarray to identify substrate pairs that formed a reaction product under the reaction conditions.

20. A method of identifying new chemical reactions, the method comprising steps of:
- providing one or more DNA templates,
    - wherein one or more templates have two substrates associated therewith;
    - wherein a portion of the sequence of the DNA template identifies the substrates associated with the template; and
    - wherein one of the two substrates and biotin is linked to the template through a cleavable linker;
  - subjecting the templates with substrates attached thereto to reaction conditions,
    - whereby a covalent bond is formed between the substrates associated with the template;
    - cleaving the linker;
    - selecting substrate pairs that reacted to form a covalent bond between the substrates using streptavidin; and
    - identifying the selected substrate pairs and reaction conditions.
21. The method of claim 20, wherein the step of identifying comprises:
- amplifying the DNA template associated the identified reaction product using PCR; and
  - sequencing the amplified DNA.
22. The method of claim 20, wherein the step of identifying comprises using a microarray to identify the selected substrate pairs that reacted.
23. A template oligonucleotide for reaction discovery comprising:
- an oligonucleotide with sequences identifying substrates attached thereto;
  - two substrates; and
  - a cleavable linker with an affinity reagent.



24. The template of claim 23, wherein the oligonucleotide is DNA.
25. The template of claim 23, wherein the oligonucleotide further comprises primer sequences.
26. The template of claim 23, wherein the cleavable linker is a disulfide bond.
27. The template of claim 23, wherein the affinity reagent is biotin.
28. A method for preparing a template oligonucleotide for reaction discovery, the method comprising steps of:
  - providing a template oligonucleotide with a first substrate attached and with a sequence that identifies the first attached substrate;
  - hybridizing through an overhand an oligonucleotide with a sequence that identifies a second substrate;
  - polymerizing off the template oligonucleotide by primer extension to include the sequence that identifies the second substrate;
  - denaturing resulting duplex; and
  - ligating onto the template oligonucleotide an oligonucleotide with the second substrate and affinity reagent linked to the oligonucleotide through a cleavable linker.
29. The method of claim 28, further comprising purifying the template oligonucleotide.
30. A kit for identifying new chemical reactions comprising one or more template oligonucleotides of claim 23 and a DNA microarray.
31. The kit of claim 30 further comprising any of the reagents selected from the group consisting of enzymes, solvents, reagents, buffers, and streptavidin beads.

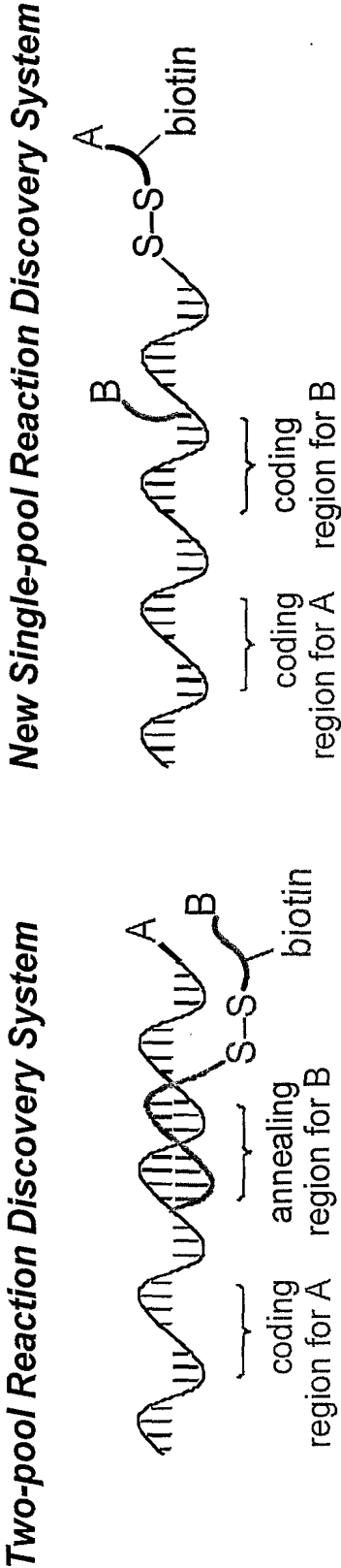


Figure 1



# Analysis of DNA sequences Encoding Bond-Forming Reactions

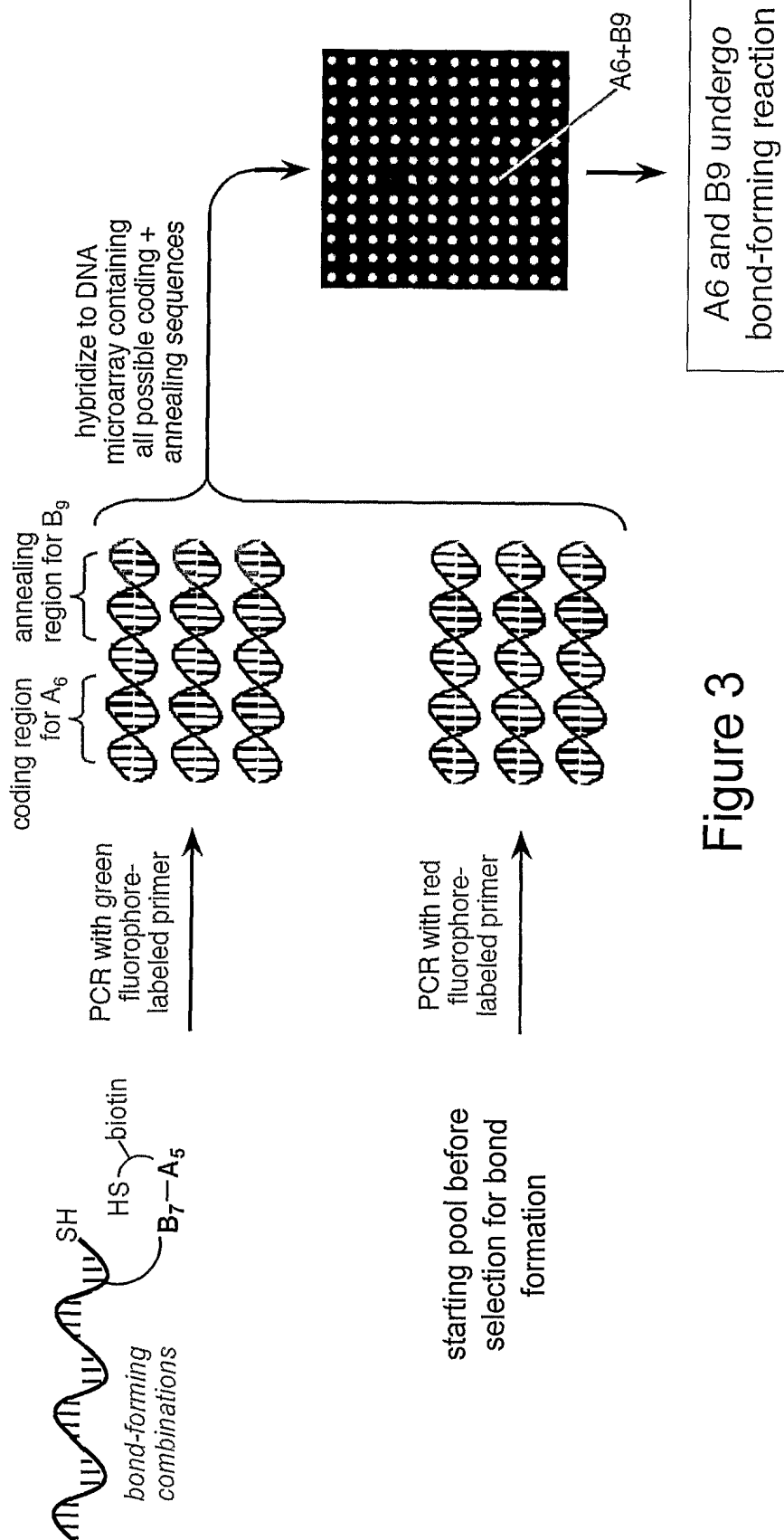
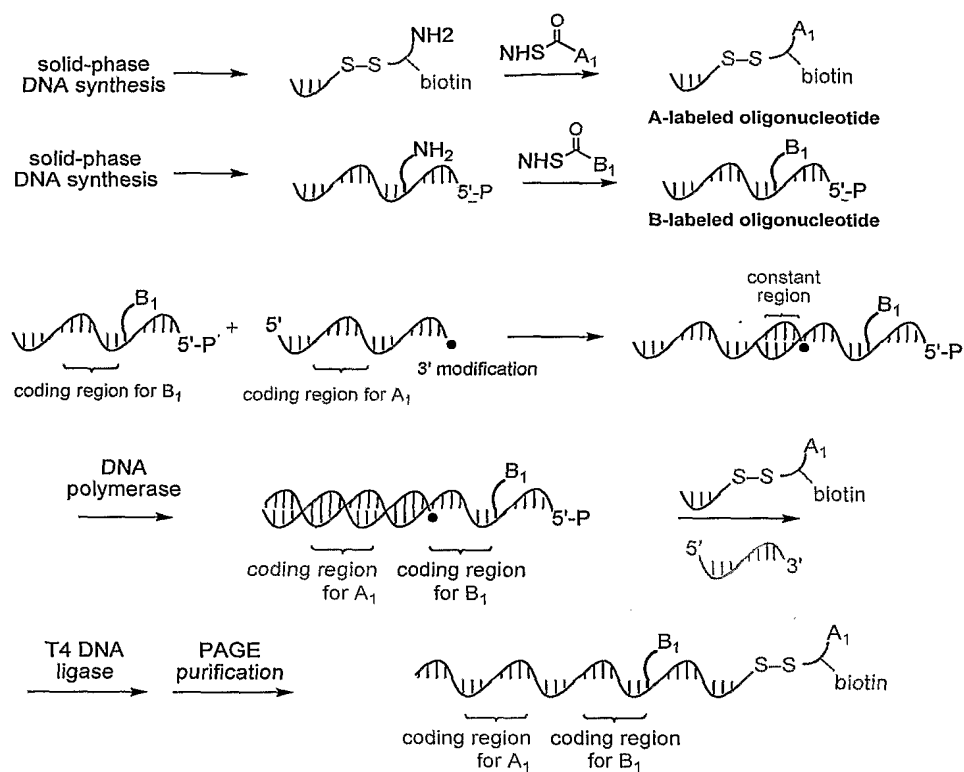


Figure 3

**Figure 4**

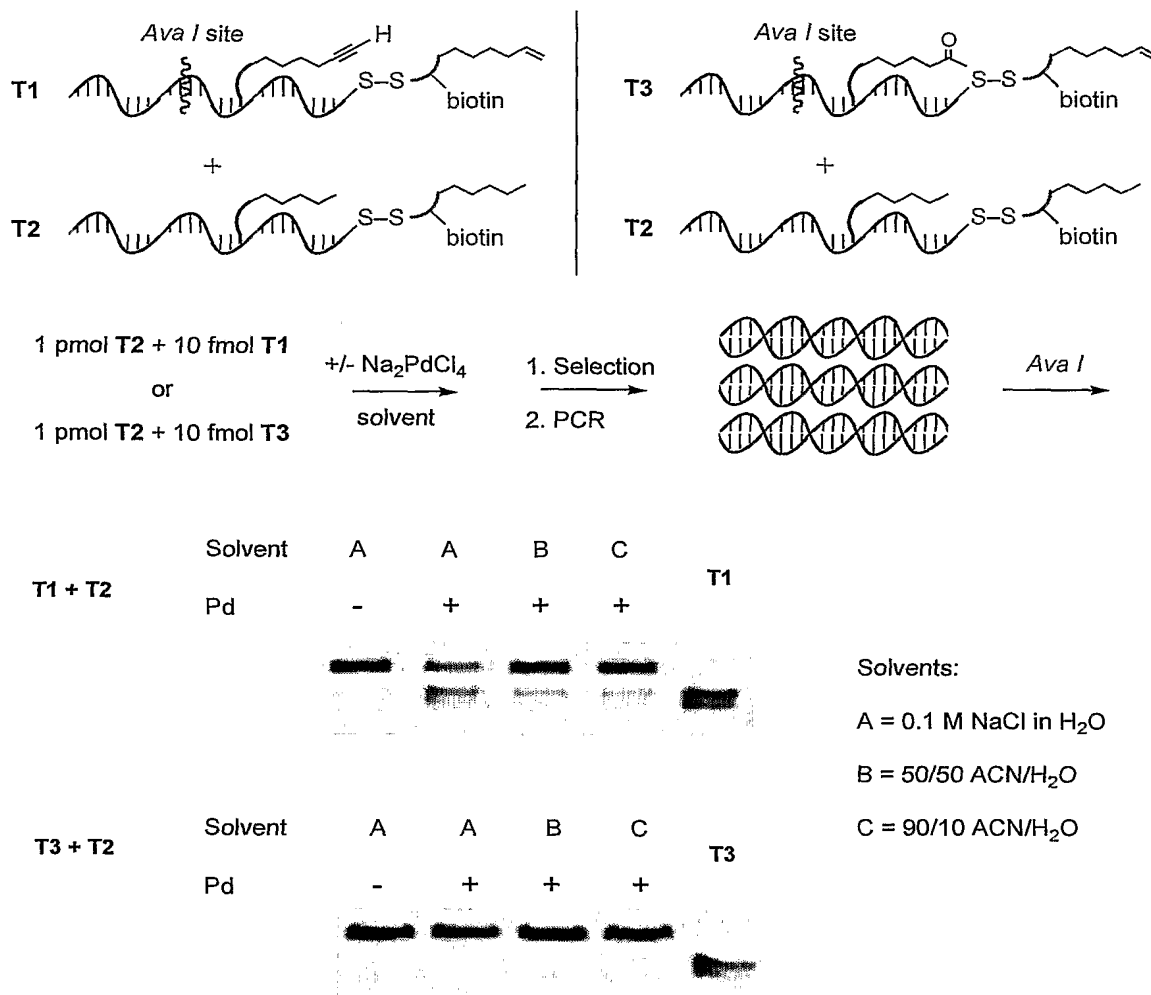
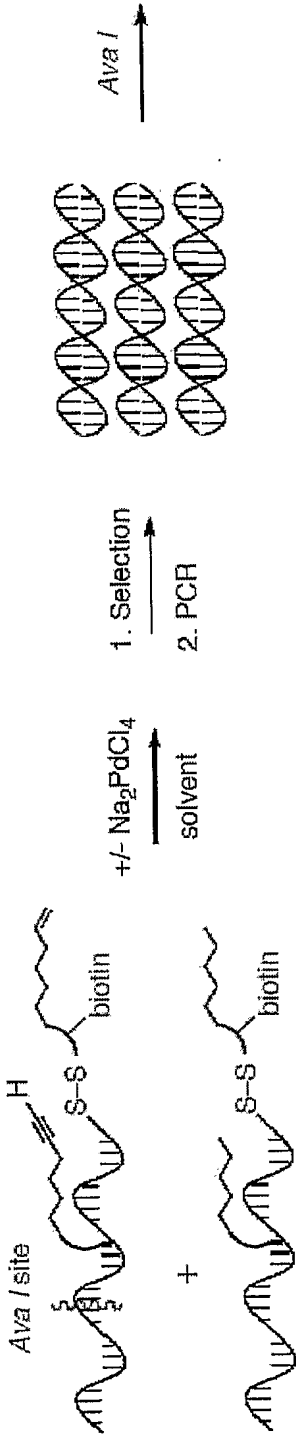


Figure 5

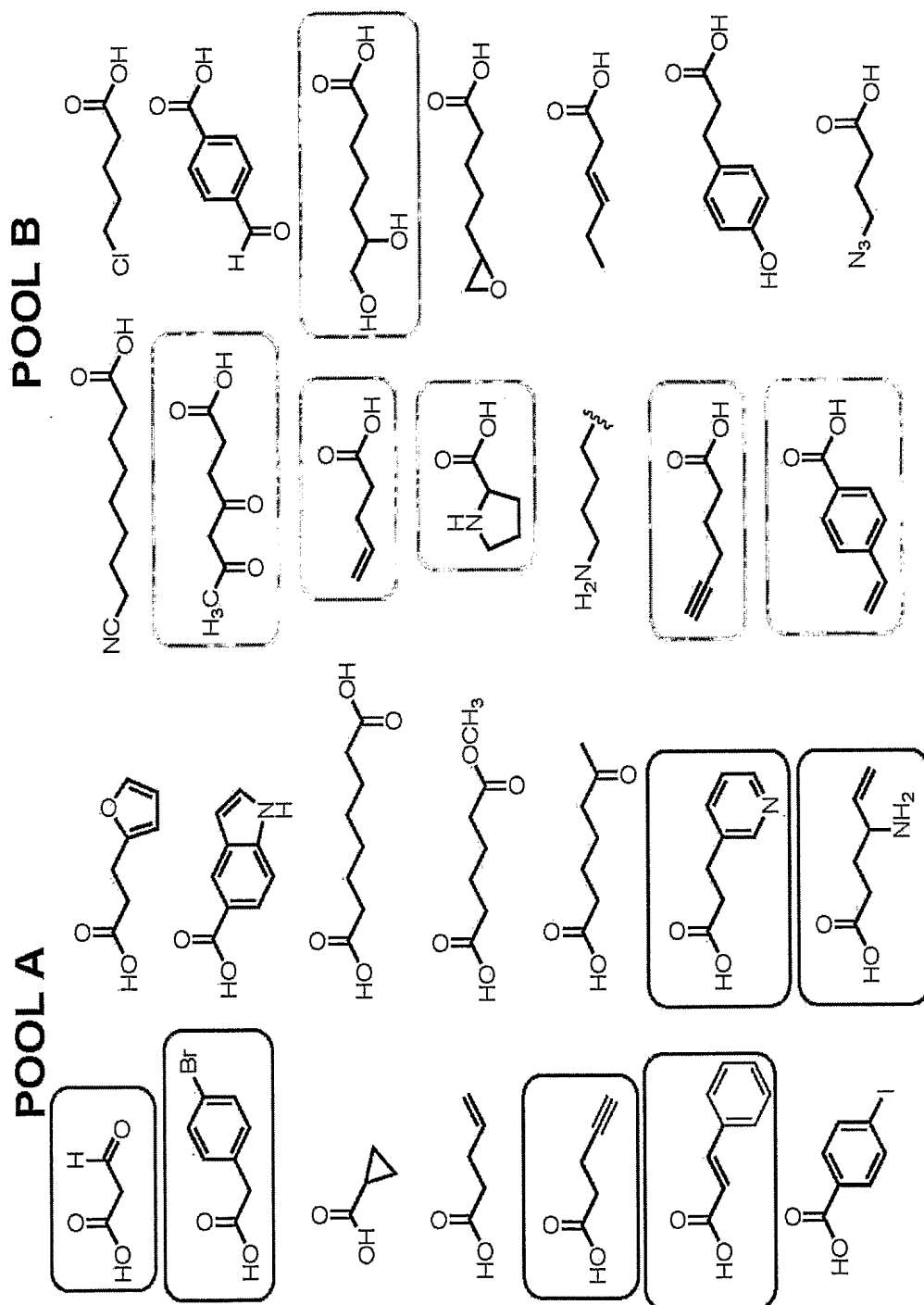
# Testing New Architecture



		Conditions:			Enrichments:	
		(A) T1 control			n/a	
Pd	A	-	-	-	(C) 1:50 (T1:T2)	>100 fold
	B	-	-	-	(B) 1:100 (T1:T2)	~100 fold
		no Ava1 cleavage in the Pd free control selections				

Figure 6

# Substrate Structures in New Matrix



**Figure 7**



# Matrix Assembly Strategy

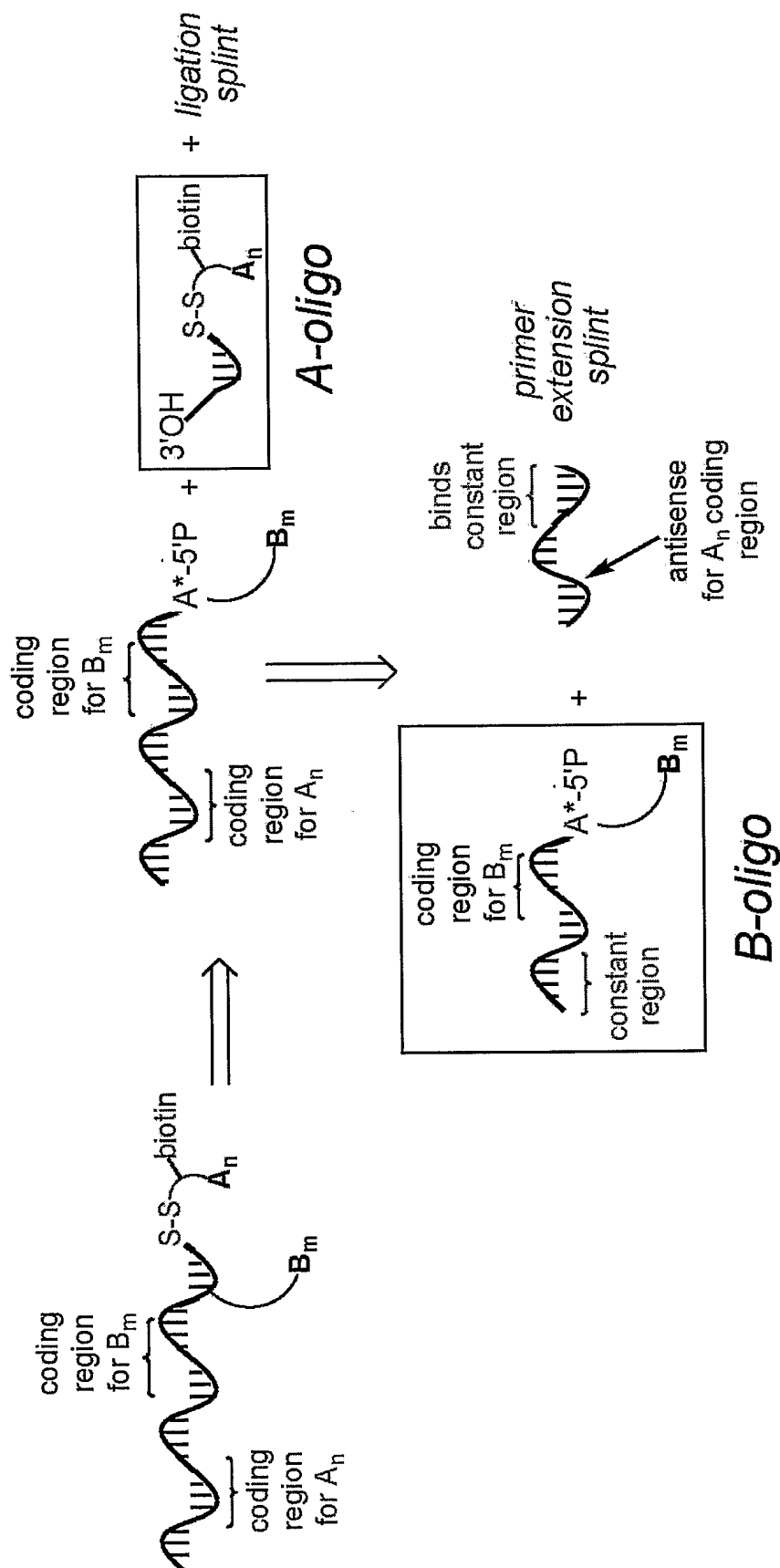
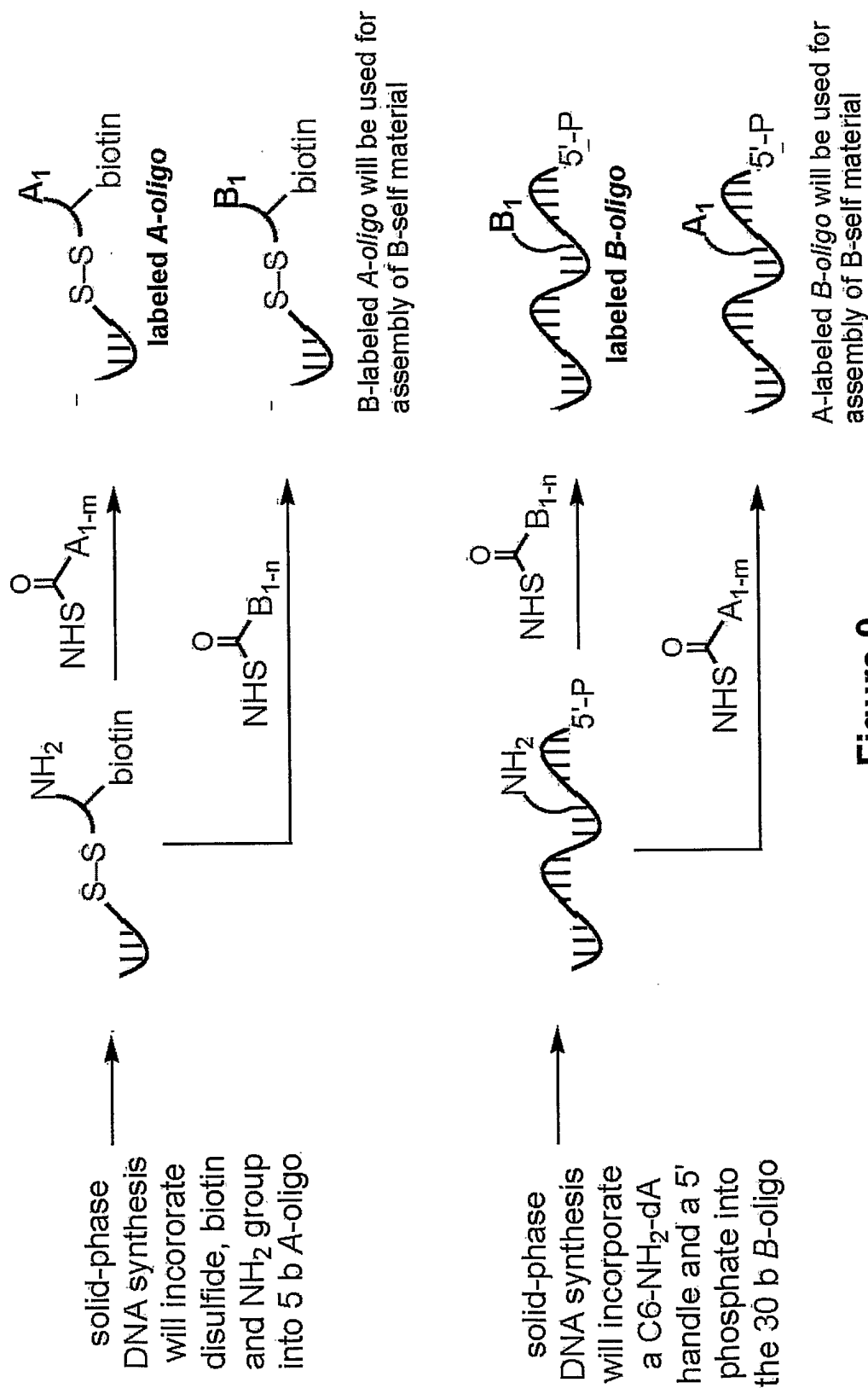


Figure 8

# Component Preparation



# Labeled Components I

Oligo ID	DNA Sequence (5'-->3')	1/epsilon	mass (no 5'A*p)	expected mass (with 5'A*p)	w/ SM	MALDI data
generalized	pA*TCCGCAG-(ANx)-CACACACC					
B-AN1-8b	pA*TCCGCAGAACTTCCTCTCGGACACACACC	3.172	9056.9	9562.26	119.03	9679
B-AN2-8b	pA*TCCGCAGACGCGATGTTTCGACCACACACC	3.122	9096.9	9602.26	133.03	9741
B-AN3-8b	pA*TCCGCAGAGCGTTATGTTCCGACACACACC	3.075	9137	9642.36	145.09	9789
B-AN4-8b	pA*TCCGCAGACATGAGCCCCACTACACACACC	3.087	9034.9	9540.26	127.08	9670
B-AN5-8b	pA*TCCGCAGCCCACTGTTACTAGGGCACACACC	3.122	9096.9	9602.26	97.07	9701
B-AN6-8b	pA*TCCGCAGCGTGTCTGAGGAGAAACACACACC	2.968	9186	9691.36	149.06	9841
B-AN7-8b	pA*TCCGCAGAGGCCCTCTTAGACCCACACACC	3.172	9056.9	9562.26	112.05	9675
B-AN8-8b	pA*TCCGCAGCTTAGTTTCGCGCACCACACACC	3.241	9047.9	9553.26	152.11	9704
B-AN9-8b	pA*TCCGCAGGAGGGTGATGCTATGTCACACACC	2.984	9217	9722.36	157.05	9870
B-AN10-8b	pA*TCCGCAGGCTGGACTAGCTACACACACACC	3.058	9105.9	9611.26	83.05	9703
B-AN11-8b	pA*TCCGCAGTAGACACGCGATGTGCCACACACC	2.983	9105.9	9611.26	115.08	9728
B-AN12-8b	pA*TCCGCAGTCGGACTTTGATGGCCACACACC	3.141	9128	9633.36	0	9627
B-AN13-8b	pA*TCCGCAGACCGAAGGGCAATACACACACC	2.935	9124	9629.36	109.06	9735
B-AN14-8b	pA*TCCGCAGGTAAGCCCTTGGTATCACACACC	3.109	9112	9617.36	116.06	9725

B-oligos labeled with pool B material B1-B14 Figure 10

# Figure 11

## Labeled Components II

Oligo ID	DNA Sequence (5'-->3')	1/epsilon	mass (no 5'A*p)	expected mass (with 5'A*p)	w/ SM	MALDI data
generalized	pA*TCCGCAG-(ANx)-CACACACC					
B-ANself-8b A1	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	99.08 9710.34	9711
B-ANself-8b A2	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	196.96 9808.22	9813
B-ANself-8b A3	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	69.03 9680.29	9690
B-ANself-8b A4	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	83.05 9694.31	9686
B-ANself-8b A5	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	109.06 9720.32	9717
B-ANself-8b A6	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	119.05 9730.31	9727
B-ANself-8b A7	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	230.93 9842.19	9840
B-ANself-8b A8	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	123.04 9734.3	9729
B-ANself-8b A9	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	144.04 9755.3	9746
B-ANself-8b A10	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	157.09 9768.35	9763
B-ANself-8b A11	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	143.07 9754.33	9754
B-ANself-8b A12	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	127.08 9738.34	9736
B-ANself-8b A13	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	134.06 9745.32	9743
B-ANself-8b A14	pA*TCCGCAGTTGATGACCAGGCCACACACACC	3.058	9105.9	9611.26	112.08 9723.34	9725

B-oligos labeled with pool A material A1-A14 (for pool A<sub>self</sub>)

# Labeled Components III

Oligo ID	DNA Sequence (5'-->3')	1/epsilon mass (no 5'link)	expected mass (w/ 5'linker)	w/ SM	MALDI data
A1oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	99.08
A2oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	2521.46
A3oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	196.96
A4oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	2619.34
A5oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	69.03
A6oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	2491.41
A7oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	83.05
A8oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	2505.43
A9oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	109.06
A10oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	2531.44
A11oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	119.05
A12oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	2541.43
A13oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	230.93
A14oligo	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	2653.31
					123.04
					2545.42
					2547.3
					144.04
					2566.42
					157.09
					2579.47
					143.07
					2565.45
					127.08
					2549.46
					134.06
					2556.44
					112.08
					2534.46
					2532.3

A-oligos labeled with pool A material

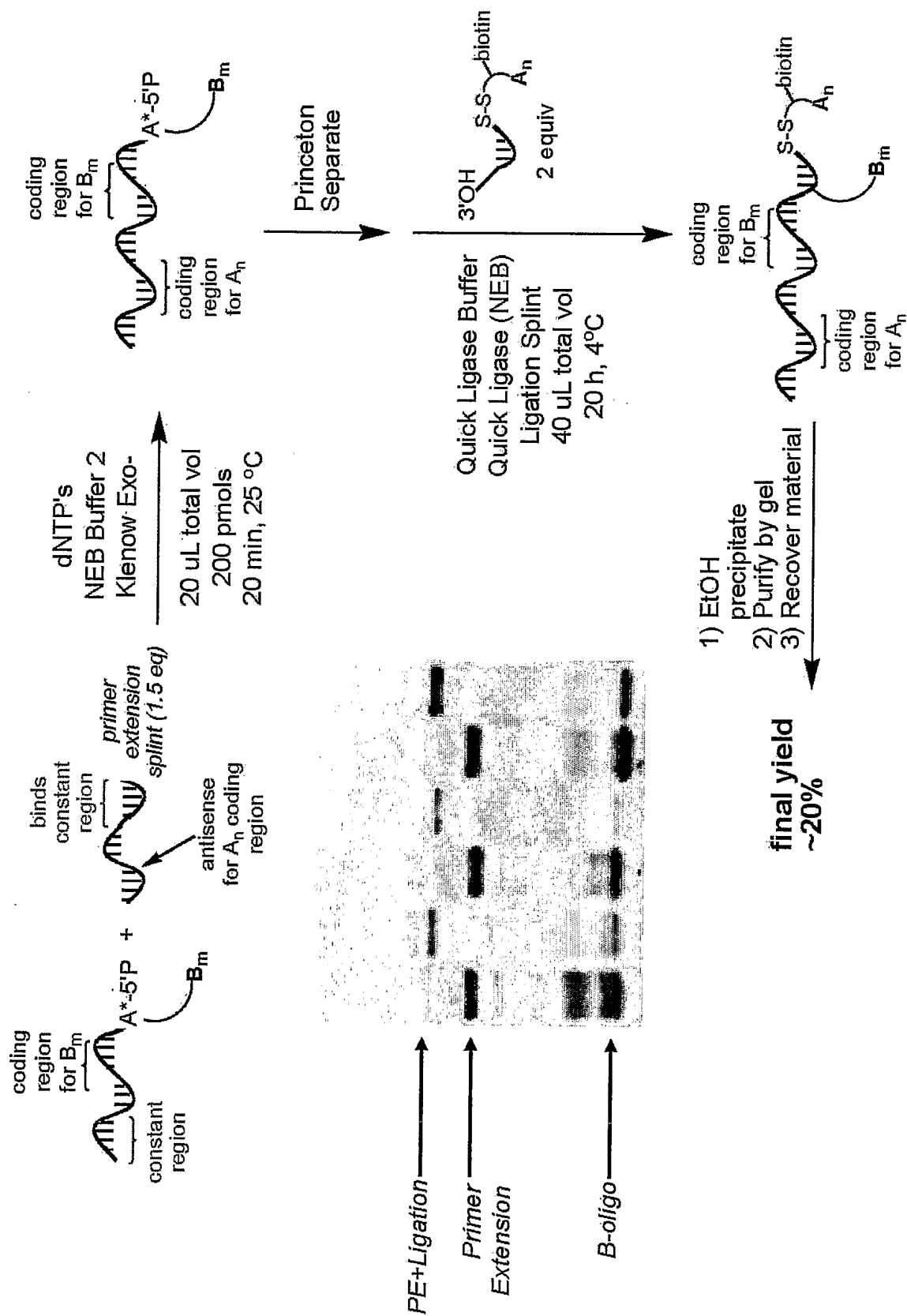
Figure 12

Figure 13  
Labeled Components IV

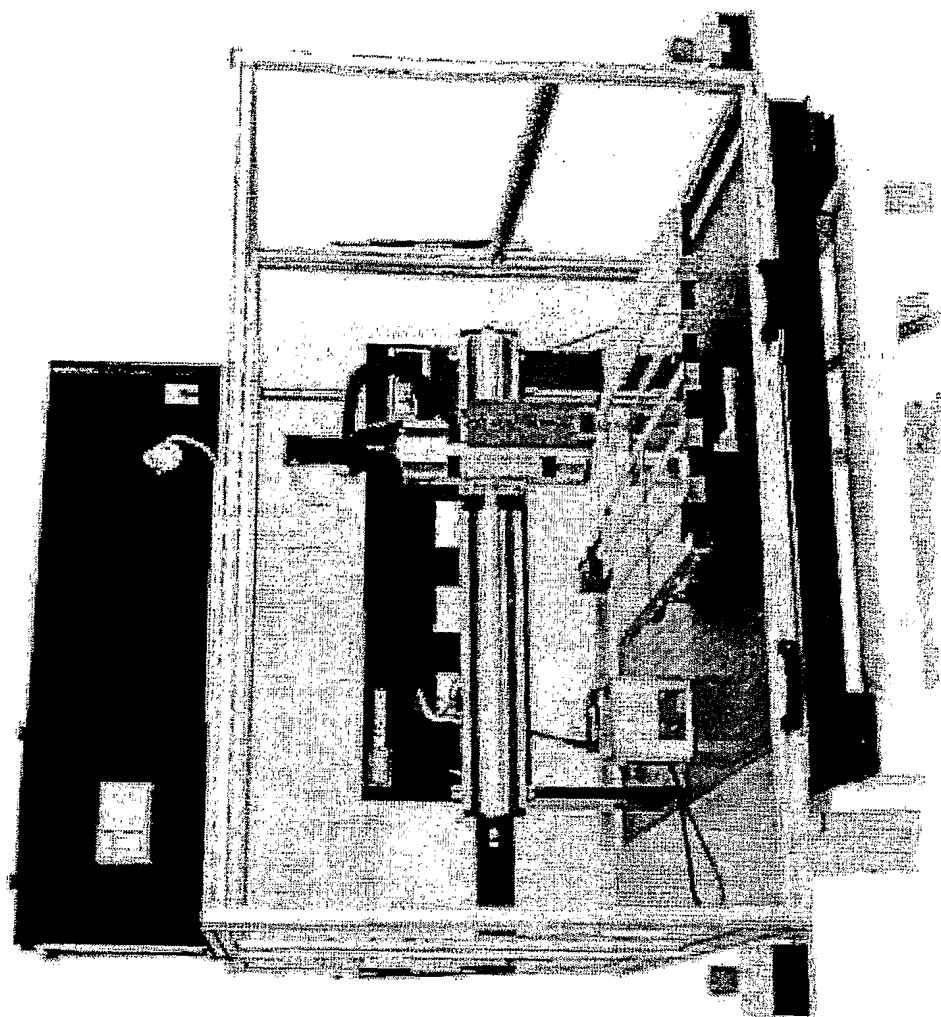
Oligo ID	DNA Sequence (5'-->3')	1/epsilon mass (no 5'link)	expected mass (w/ 5'linker)	w/ SM	MALDI data
AselFB1	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	119.03 2541.41 2540
AselFB2	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	133.03 2555.41 2554.7
AselFB3	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	145.09 2567.47 2567
AselFB4	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	127.08 2549.46 2547
AselFB5	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	97.07 2519.45 2518
AselFB6	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	149.06 2571.44 2570.2
AselFB7	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	112.05 2534.43 2533.7
AselFB8	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	152.11 2574.49 2574
AselFB9	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	157.05 2579.43 2580
AselFB10	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	83.05 2505.43 2502
AselFB11	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	115.08 2537.46 2532
AselFB12	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	0 2422.38 2423
AselFB13	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	109.06 2531.44 2531.3
AselFB14	NH2--biotin--C6S-SC6--TTGAT	19.08	1493.1	2422.38	116.06 2538.44 2540

A-oligos labeled with pool B material (for B<sub>self</sub> pool)

Figure 14  
Matrix Assembly Trials



## Figure 15 Microarray Technology



DNA Microarrays have been printed with amino-terminated oligos on aldehyde silane slides (Nextion slide-AL).

Features correspond to each  $A_m \times B_n$  encoding oligo ( $14 \times 14 = 196$ ) as well as each homocoupling combination (28) and an internal standard (1)

$$196 + 28 + 1 = 225 = 15 \times 15$$



# Validating Known Reactions

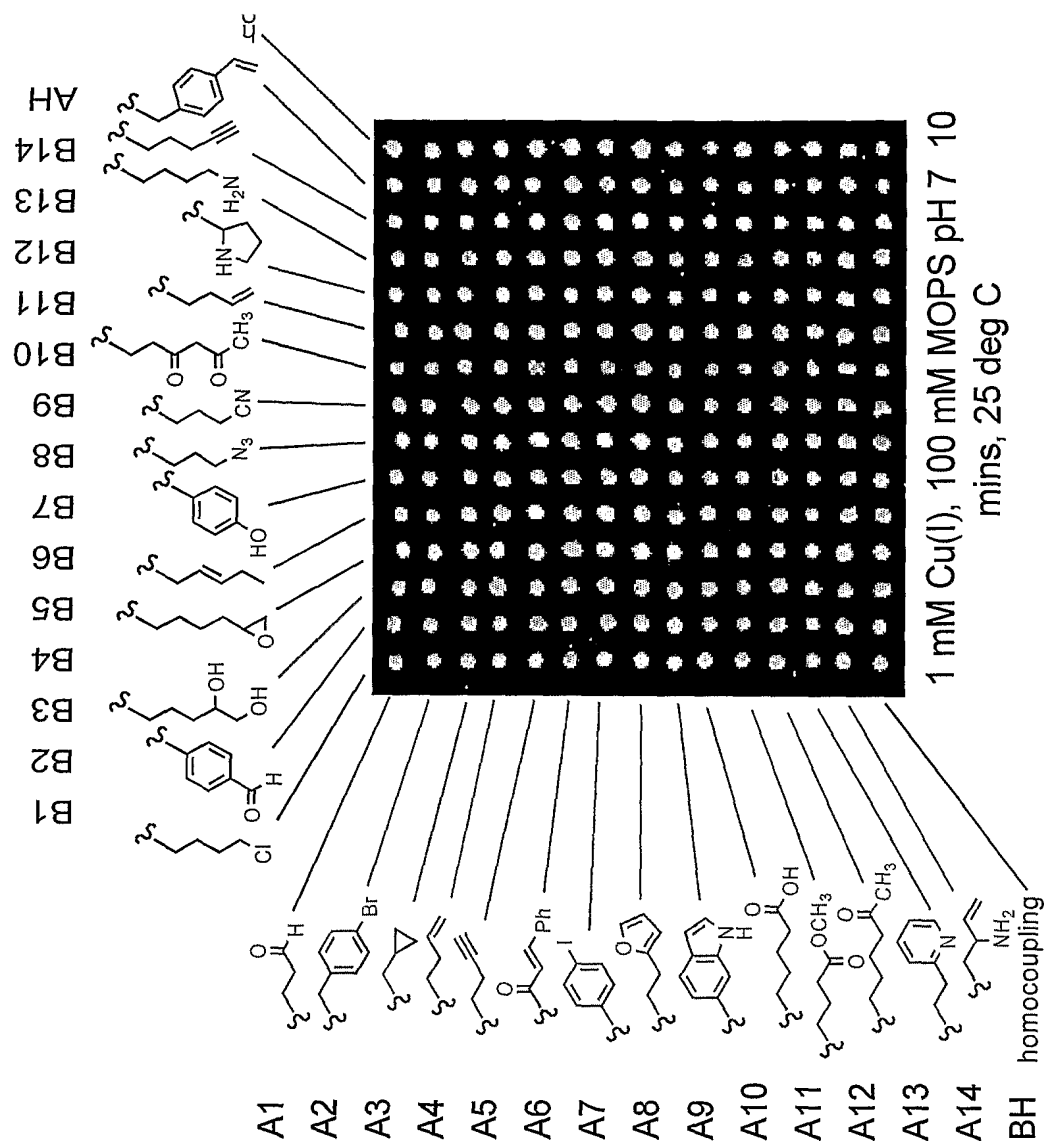
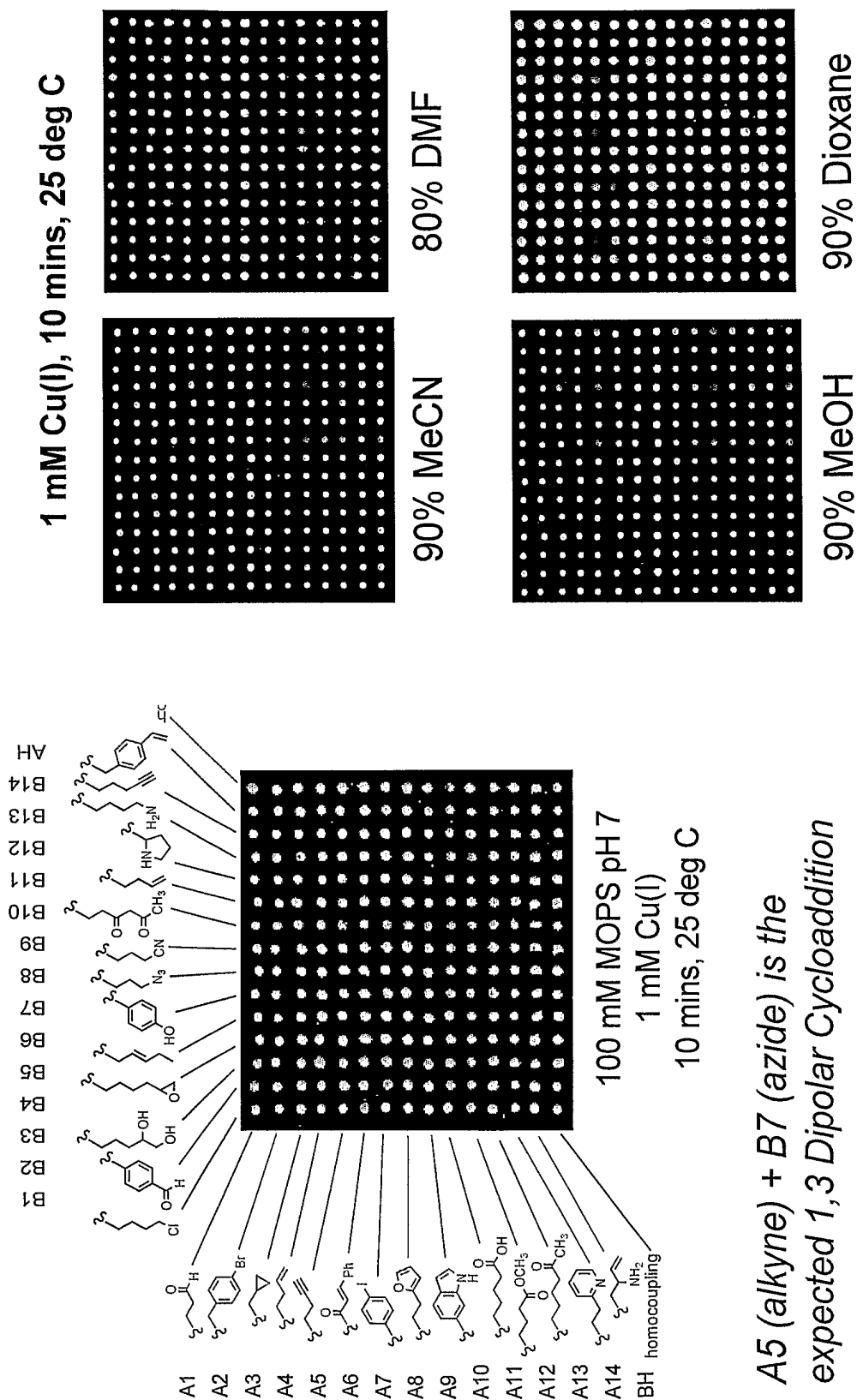


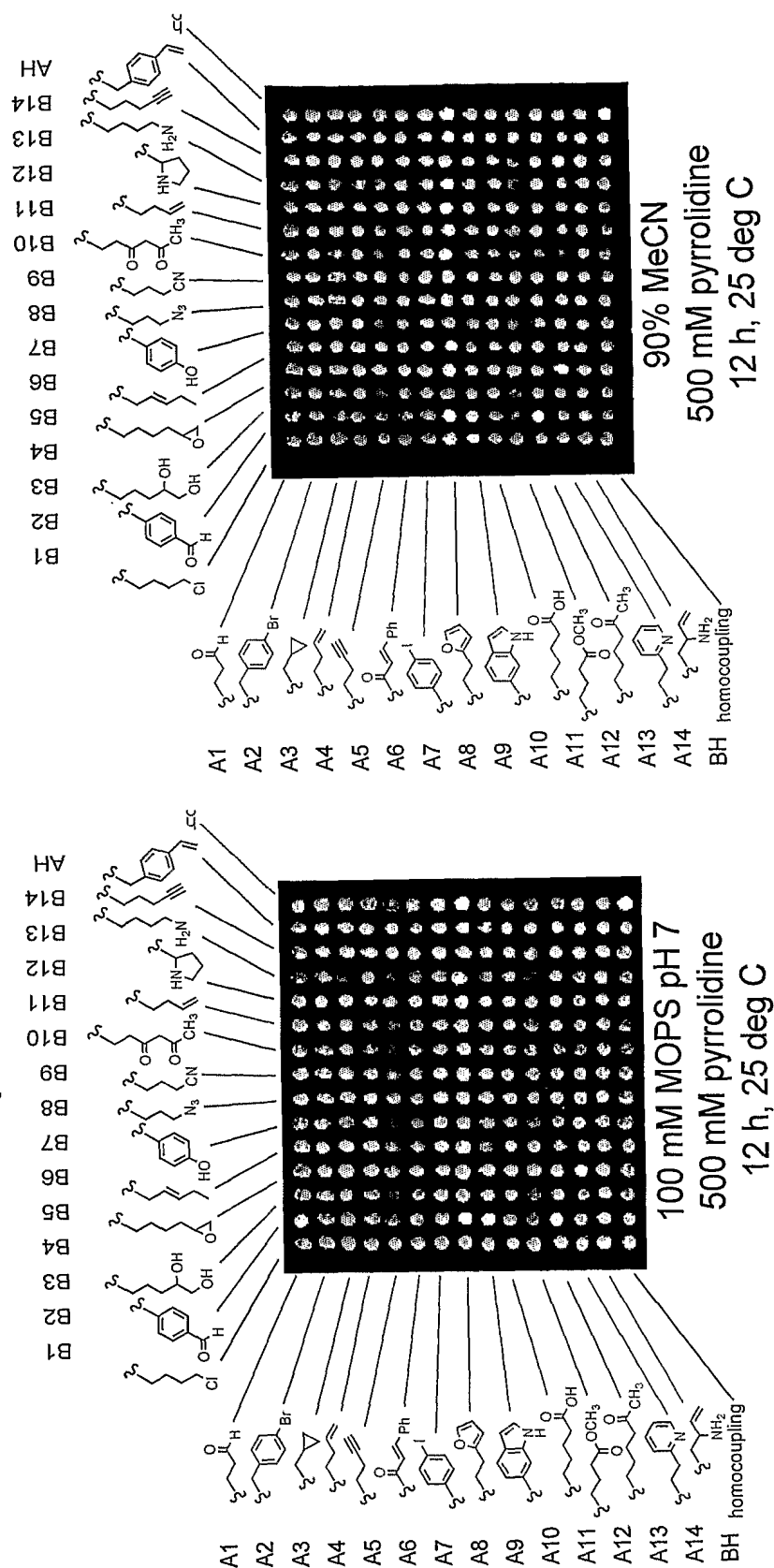
Figure 16

# Figure 17

## Cu(I) Reactivity Across Solvents



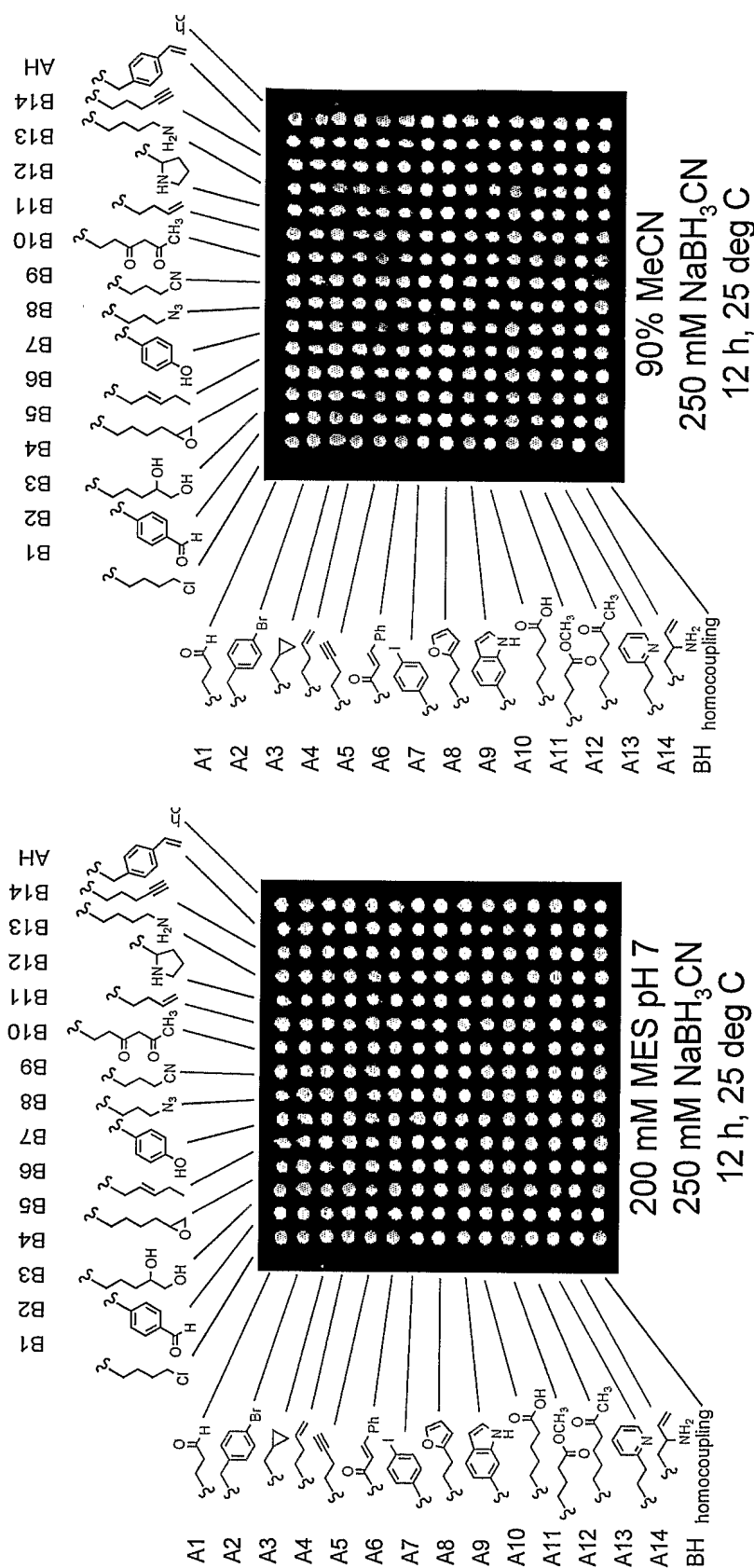
# More Proof of Principle (Enamine Chemistry)



Pyrrolidine is expected to mediate an enamine aldol reaction between the ketone (A12) and aryl aldehyde (B2)

**Figure 18**

# Still More Proof of Principle (Reductive Amination)



NaBH<sub>3</sub>CN is expected to mediate a reductive amination between the allyl amine (A14) and aryl aldehyde (B2)

**Figure 19**

# Reactivity Across Temperatures

1 mM Pd(II) in MOPS pH 7, 20 min

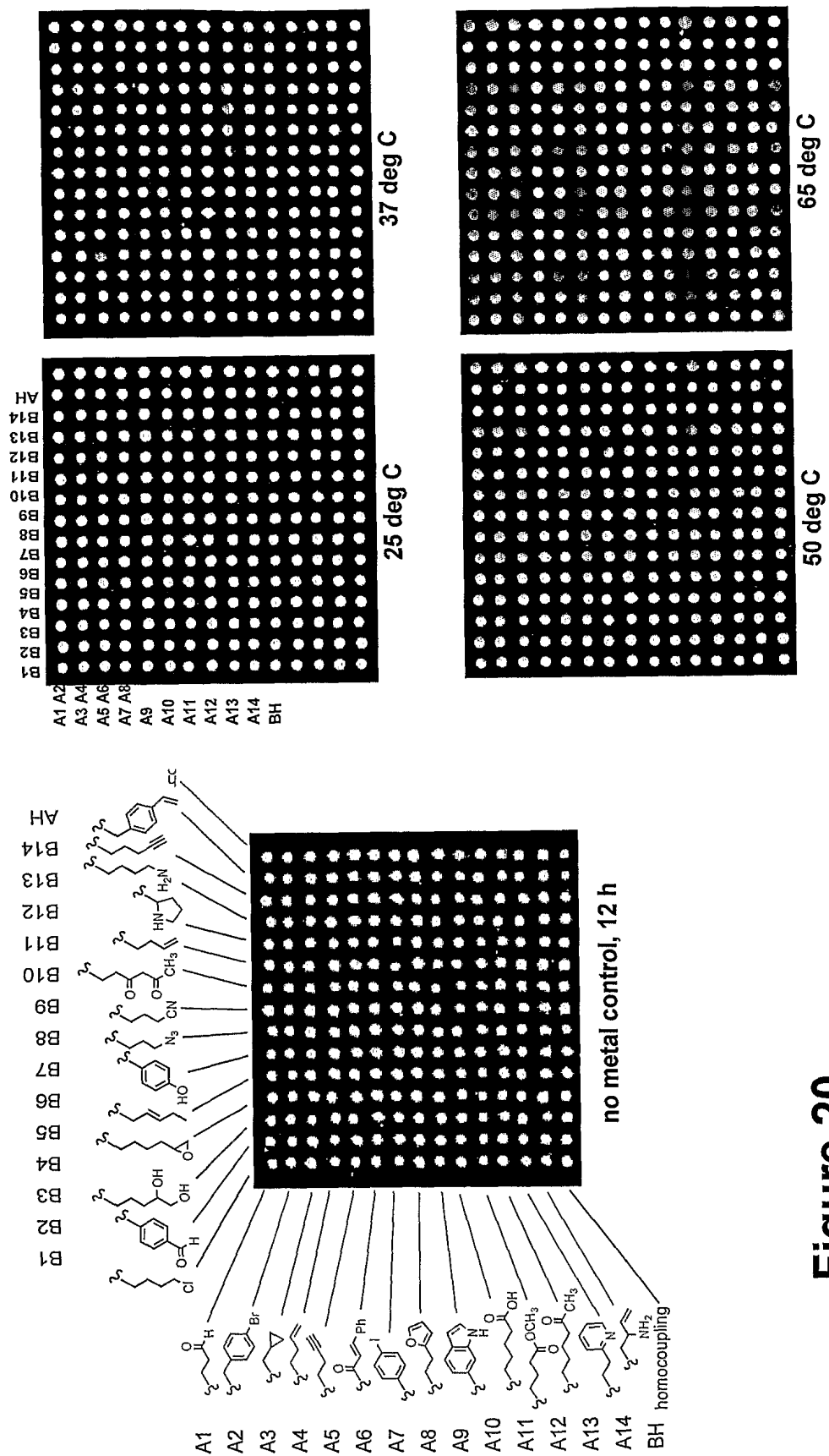


Figure 20

# Reactivity Across Temperatures

1 mM Pd(II) in 90% MeCN, 20 min

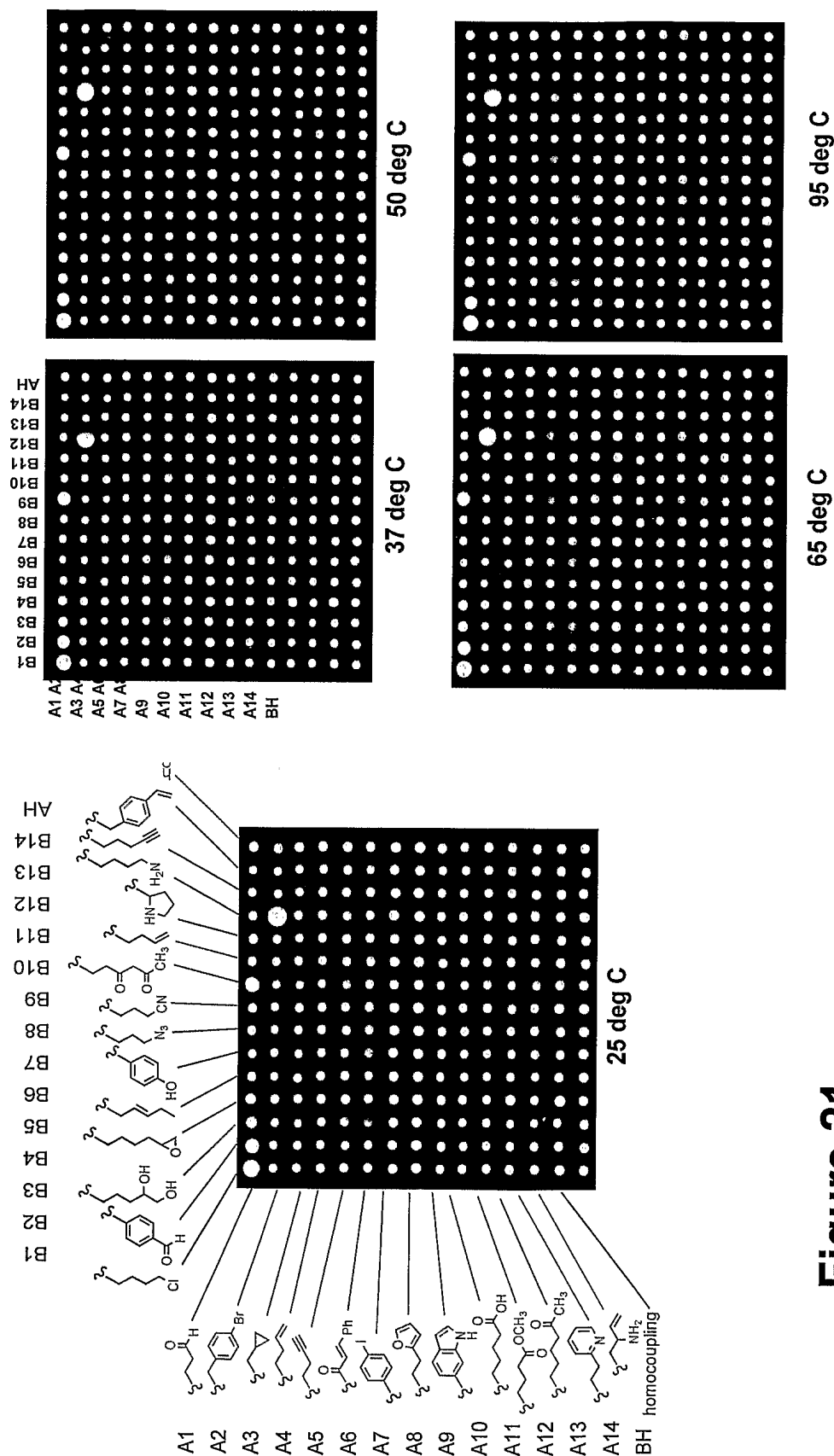


Figure 21

# Reactivity Across Temperatures

1 mM Pd(II) in 90% DMSO, 20 min

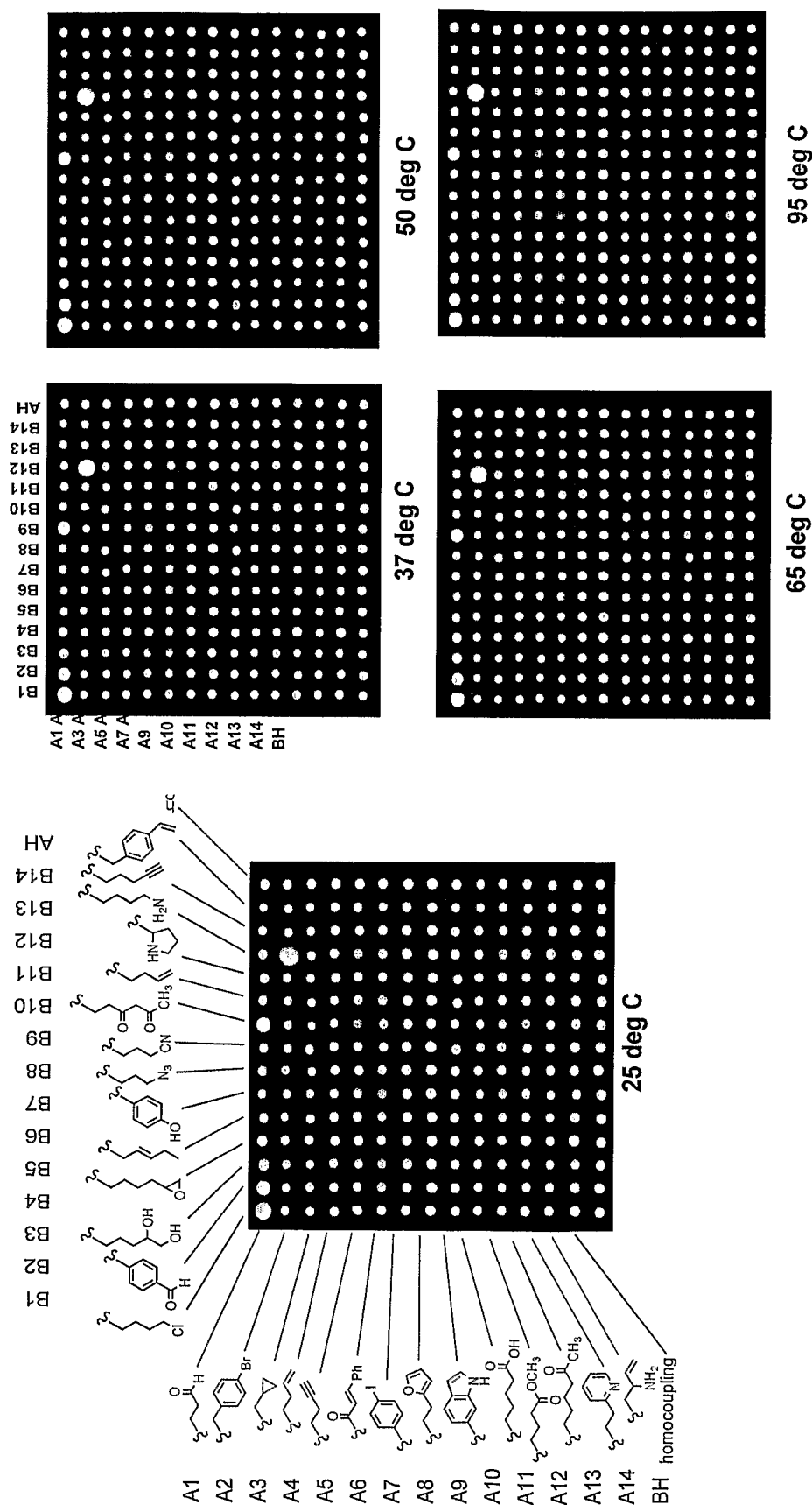


Figure 22